Molecular biological characterisation of resectable pancreatic ductal adenocarcinoma – identifying a signature of responsiveness to erlotinib

DISSERTATION

zur Erlangung des akademischen Grades

Doctor rerum naturalium

(Dr. rer. nat.)

eingereicht an der

Lebenswissenschaftlichen Fakultät der Humboldt-Universität zu Berlin

Von

Kaja Hoyer, M.Sc.

Präsidentin der Humboldt-Universität zu Berlin

Prof. Dr.-Ing. Dr. Sabine Kunst

Dekan der Lebenswissenschaftlichen Fakultät

der Humboldt-Universität zu Berlin

Prof. Dr. Dr. Christian Ulrichs

Gutachter/innen

1. Prof. Dr. Nils Blüthgen

2. Prof. Dr. med. Frederik Damm

3. PD Dr. med. P. Ivanyi

Tag der mündlichen Prüfung: 08.09.2021

The following work was carried out in the research group of Frederik Damm between December 2015 and June 2020. Partial results of the presented work have been accepted for publication in EbioMedicine[1].

1. Gutachter: Prof. Dr. Nils Blüthgen
IRI Life Sciences
Charite - Universitätsmedizin Berlin

2. Gutachter: Prof. Dr. med. Frederik Damm
Klinik für Hämatologie, Onkologie und Tumorimmunologie
Charité-Universitätsmedizin Berlin

3. Gutachter: PD Dr. med. P. Ivanyi
Klinik für Hämatologie, Onkologie und Stammzelltransplantation
Medizinische Hochschule Hannover

Abstract

Our knowledge of the genetic drivers of pancreatic ductal adenocarcinoma (PDAC) has increased rapidly in recent decades. Nevertheless, its prognosis remains dismal. In contrast to other cancer entities, PDAC patients have not benefited from recent improvements in precision medicine. To address this gap, I embarked on a comprehensive molecular study to identify predictive biomarkers and refine risk stratification. I performed targeted sequencing and targeted RNA expression analysis of 293 R0-resected patients (CO-2016 cohort) from a multicenter phase III trial comparing adjuvant chemotherapy of gemcitabine with or without erlotinib. Patients were clustered using non-negative matrix factorization (NMF) based on their single nucleotide variant (SNV) and copy number alteration (CNA) statuses. Overall (OS) and disease-free survival (DFS) were analysed with the multivariate cox hazard and log rank tests. Finally, using a method based on CRISPR/Cas, findings from the patient cohort where modeled *in vitro* to assess their biological backgrounds.

A total of 1,086 SNVs and 4,157 CNAs were found with at least one genetic alteration in 99% of all patients, and an average of 18 aberrations per patient were found. In line with previous reports, *KRAS*, *TP53*, *CDKN2A*, and *SMAD4* were the most frequently affected genes, detected in 63–93 % of cases. To identify the complex interplay of genetic aberrations, I used NMF and subsequent hierarchical clustering to define five patient subgroups. These groups differed in SNV and CNA patterns, transition/transversion rates, and gene expression profiles as well as in patient OS and DFS. This led to the identification of two distinct biological profiles underlying PDAC development and a potential biomarker for erlotinib treatment response. Patients in the cluster with a significant accumulation of *SMAD4* mutations showed increased sensitivity toward erlotinib treatment. The effect was even more apparent when comparing treatment arms (Gemcitabine vs Gemcitabine + Erlotinib) in all patients with either an SNV or CNA in *SMAD4* (HR=0.52; test for interaction, p=0.026). Subsequent integration of differential gene expression analysis established *SMAD4* alterations and low *MAPK9* expression (n=91) as a predictive biomarker for longer DFS (HR=0.49; test for interaction, p=0.02) and OS (HR=0.32; test for interaction, p=0.001) in erlotinib treated patients. Modeling of $SMAD4^{alt}$ $MAPK9^{low}$ status *in vitro* showed that the effect is not based on increased erlotinib toxicity. Finally, I proposed a genetic risk score for prognostic evaluation of newly diagnosed R0-resected PDAC patients. The score was based on two clinical (N-stage and grading) and three molecular variables (gene expression levels of *DDIT4*, *HIST1H3H,* and *ITGA3).* Median overall survival (OS) from patients in the low (n=63), intermediate (n=115) and high-risk (n=51) groups were 13, 26, and 50 months respectively and could be validated in an independent patient cohort.

In this thesis, I identified five biologically distinct patient subgroups with different actionable lesions that may serve for refined PDAC classification and tailored treatment approaches. In addition, I proposed a genetic risk score that may guide adjuvant chemotherapy intensity and ultimately improve patient outcomes. Finally, I determined an *SMAD4* alteration and low *MAPK9* expression as a potential predictive biomarker for erlotinib sensitivity.

Zusammenfassung

Durch die Weiterentwicklung moderner Sequenziertechnolgien hat sich innerhalb des letzten Jahrzehnts das Wissen der genetischen Veränderungen des duktalen Pankreas-Adenokarzinom (PDAC) vervielfacht. Im Vergleich zu anderen Krebsentitäten, konnten Patienten mit PDAC jedoch bisher kaum von den Therapieerfolgen der Präzisionsmedizin profitieren. Um diese Problematik zu adressieren, habe ich eine umfassende molekularbiologische Studie durchgeführt, in deren Rahmen prädiktive Biomarker identifiziert und die Risikostratifizierung der Patienten verfeinert werden sollte. Mittels gen-spezifischer Sequenzierung und gezielter RNA-Expressionsanalyse wurden 293 R0-resezierte Patienten (CO-2016 cohort) aus einer multizentrischen Phase-III-Studie untersucht. Ziel der klinischen Studie war der Vergleich von adjuvanter Chemotherapie mit Gemcitabin entweder mit oder ohne Zusatz von Erlotinib. Die Patienten wurden unter Verwendung einer nicht-negativen Matrixfaktorisierung (NMF) basierend auf ihren Einzelnukleotidvarianten (SNV) und ihren Kopienzahlveränderungen (CNA) gruppiert und auf klinische und molekularbiologische Unterschiede untersucht. Um die biologischen Hintergründe der identifizierten genetischen Besonderheiten zu verstehen, wurden Zelllinien genetisch modifiziert und *in vitro* modelliert.

Es wurden 1086 SNVs und 4157 CNAs identifiziert. Dabei wiesen 99% aller Patienten mindestens eine genetische Veränderung auf, mit durchschnittlich 18 Aberrationen pro Patient. In Übereinstimmung mit früheren Berichten waren *KRAS*, *TP53*, *CDKN2A* und *SMAD4* die am häufigsten betroffenen Gene. Alterationen in diesen Genen konnten in 63-93 % der Fälle nachgewiesen werden. Um im komplexen Zusammenspiel dieser genetischeren Aberrationen biologisch relevante Muster zu identifizieren, habe ich sie mittels NMF gruppiert. Basierend darauf konnte ich durch hierarchisches Clustern fünf Patientenuntergruppen definieren. Diese Gruppen unterschieden sich in SNV- und CNA-Häufigkeit, hatten unterschiedliche Genexpressionsprofile sowie signifikant längeres bzw. kürzeres OS und DFS. So konnte ich nicht nur unterschiedliche biologische Profile identifizieren die auf unterschiedliche Art die PDAC Entwicklung ermöglichen, sondern zudem einen ersten Hinweis auf einen potenziellen Biomarker für das Ansprechen auf die Erlotinib-Behandlung finden. Patienten aus einem Cluster mit signifikanter Anhäufung von *SMAD4* Mutationen, zeigten eine erhöhte Erlotinib Sensitivität. Dieser Effekt wurde sogar noch verstärkt beim Vergleich der beiden Behandlungsarme (Gemcitabine vs Gemcitabine + Erlotinib) in allen Patienten mit einem SNV oder CNA in *SMAD4* (HR=0,52; Interaktionstest, p=0,026). Eine anschließende Verknüpfung dieses Ergebnisses mit den Daten der differentiellen Genexpressionsanalyse ergab, dass die Kombination von *SMAD4*-Veränderungen und niedrigem *MAPK9*-Expressionlevel (n = 91) ein potentieller prädiktiver Biomarker für adjuvante Behandlung in R0-resezierten PDAC Patienten ist. $SMAD4^{alt}$ $MAPK9^{low}$ Patienten die mit Erlotinib behandelt

wurden wiesen ein längeres DFS (HR = 0,49; Interaktionstest, p = 0,02) und OS (HR = 0,32; Interaktionstest, p = 0,001) auf, im Vergleich zu *SMAD4*^{alt} *MAPK9*^{low} Patienten ohne Erlotinib. In einem *in vitro* Modell des Zustands (*SMAD4*^{alt} *MAPK9*^{low}) konnte ich anschließend zeigen, dass der Überlebenseffekt in der Patientenkohorte nicht auf einer erhöhten Zytotoxizität von Erlotinib gegenüber den veränderten Krebszellen beruht.

Zuletzt definierte ich basierend auf den Alterationsmustern meiner Kohorte einen genetischen Risiko-Score für die klinische Risikostratifizierung von R0-resezierten PDAC-Patienten. Der Score bestand aus zwei klinischen (N-Stage und Grading) und drei molekularen Variablen (Genexpressionsniveaus von *DDIT4*, *HIST1H3H* und *ITGA3*). Das mediane OS von Patienten in den daraus resultierenden Patientengruppen mit niedrigem (n=63), mittlerem (n=115) und hohem Risiko (n=51) unterschied sich signifikant voneinander und betrug jeweils 13, 26 bzw. 50 Monate.

Zusammenfassend lässt sich sagen, dass ich in dieser Arbeit fünf biologisch unterschiedliche Patientenuntergruppen identifiziert habe. Diese können für eine verfeinerte PDAC-Klassifizierung und für personalisierte Behandlungsansätze herangezogen werden. Darüber hinaus schlage ich einen genetischen Risiko-Score vor, der dazu beitragen kann, die richtige adjuvante Chemotherapie zu wählen und das Behandlungsergebnis zu verbessern. Schließlich postuliere ich, dass *SMAD4*-Veränderung und eine niedrige *MAPK9*-Expression als potenzielle prädiktive Biomarker für die Erlotinib-Empfindlichkeit von R0-resezierten PDAC Patienten dienen können.

Table of Content

Abbreviations

5-FU	Fluorouracil
AJCC	American Joint Committee on Cancer
Akt	Protein kinase B
Ank	Ankyrin
APC	Antigen-presenting cells
APS	Ammonium peroxydisulfate
AR	Allelic ration
ASR	Age-standardized rate
ATCC	American Type Culture Collection
BMP	Bone morphogenetic protein
BSA	Bovine serum albumin
CA19-9	Carbohydrate antigen 19-9
CAR	Chimeric antigen receptor
CNA	Copy number alteration
CO-2016 cohort	293 patients form the CONKO-005 study sequenced and analysed in the thesis
CREB	c-myc and cAMP response element-binding protein
CT	Computed tomography
DAG	Diacylglycerol
ddPCR	Digital droplet PCR
DFS	Disease-free survival
DMSO	Dimethylsulfoxid
DNA	Deoxyribonucleic acid
DSMZ	German Collection of Microorganisms and Cell Cultures
ECM	Extracellular matrix
EDTA	Ethylene diamine tetraacetic acid
EGF	Epidermal growth factors
EGFR	Epidermal growth factor receptor
EMT	Epithelial-mesenchymal transition
ERK	Extracellular signal-regulated kinase
EUS	Endoscopic ultrasound

FBS	Fetal bovine serum
FDA	Federal drug administration
FFPE	Formalin-fixed paraffin embedded
FoxO	Forkhead-box-protein
GPCR	G-protein coupled transmembrane receptor
HA	Hyaluronic acid
HNSCC	Head and neck squamous cell carcinoma
IGV	Integrative Genomics Viewer
IP3	1D-myo-inositol 1,4,5-trisphosphate
IPMN	Intraductal papillary mucinous neoplasm
JNK	c-Jun N-terminal kinase
KO	Knockout
MAPK	Mitogen-activated protein kinase
MCN	Mucinous cystic neoplasm
MEK	MAPK extracellular signal regulated kinase
mFOLFIRINOX	modified FOLFIRINOX (fluorouracil, leucovorin, irinotecan, and oxaliplatin)
MH1/2	MAD homology 1/2
MLPA	Multiplex ligation-dependent probe amplification
MSI	Microsatellite instability
mTOR	Mammalian target of rapamycin
NF-κB	Nuclear factor κB
NGS	Next generation sequencing
NMF	Non-negative matrix factorization
NSCLC	Non-small-cell lung carcinoma
OR	Odds ratio
OS	Overall survival
PanIN	Pancreatic intraepithelial neoplasia
PanNET	Pancreatic neuroendocrine tumours
PDAC	Pancreatic ductal adenocarcinoma
PDK1	3-phosphoinositide-dependent protein kinase-1
PFS	Progression-free survival

PH	Pleckstrin-homology
PIP_3	Phosphatidylinositol (3,4,5) trisphosphates
PI3K	Phosphatidylinositol-3 kinase
PLCG2	Phospholipase C Gamma 2
PSC	Pancreatic stellar cells
PTEN	Phosphatase and tensin homolog
qPCR	Quantitative real-time PCR
Raf	Rapidly accelerated fibrosarcoma
Ras	Rat sarcoma viral oncogene homolog
RD	Regulatory domain
RNA	Ribonucleic acid
RNP	Ribonucleoprotein
RTK	Receptor tyrosine kinases
sgRNA	Single guide RNA
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
TAD	Trans-activation domain
tCN	Total copy number
TEMED	N,N,N′,N′-Tetramethylethylenediamine
TGFα	Transforming growth factor alpha
TGFβ	Transforming growth factor beta
TNF	Tumour-necrosis factor
TNM	Classification of malignant tumours based in primary tumour size, lymph node involvement and metastasis status
UPD	Uniparental disomy
VAF	Variant allele frequency
ZeBanC	Central biobank of the Charité

Figure Index

Table Index

1 Introduction

1.1 Pancreatic cancer

Pancreatic cancer is one of the deadliest cancers worldwide with very limited treatment options and a prognosis that has only improved marginally in recent decades. While research into its biological background, improved diagnostics, and new therapeutic targets have increased our knowledge of the disease, late detection and early dissemination remain major treatment obstacles and make further in-depth analysis essential.

1.1.1 Clinical characteristics

1.1.1.1 Types of pancreatic cancer

While there are many subtypes of pancreatic cancer, they can all be divided into two general groups: exocrine and neuroendocrine tumours. The vast majority of pancreatic tumours (95%) occur in the exocrine system, responsible for production and distribution of digestive enzymes[2]. The most common type of exocrine pancreatic cancer is the pancreatic adenocarcinoma, which accounts for about 85–90% of all pancreatic cancer diagnoses and commonly arises from the pancreatic ducts (PDAC)[3,4]. Autopsy data have shown that approximately 60–70% of adenocarcinomas are located anatomically in the head, 5–10% in the body, and 10–15% in the tail of the pancreas [5] (Figure 1.1). It is also possible in rare cases (1–2% of exocrine tumours) for adenocarcinoma to develop from the pancreatic enzyme, creating acini. These tumours exhibit very similar symptoms to PDAC arising from ductal cells, but have a slightly better prognosis[6]. A wide array of other rare exocrine tumors include: adenosquamous carcinoma, colloid carcinoma, giant cell tumour, hepatoid carcinoma and undifferentiated carcinoma. This wide array of exocrine tumours is only diagnosed in <1% of all pancreatic cancer patients and exhibits overlapping histopathological characteristics which can make diagnosis and subsequent characterisation challenging[3].

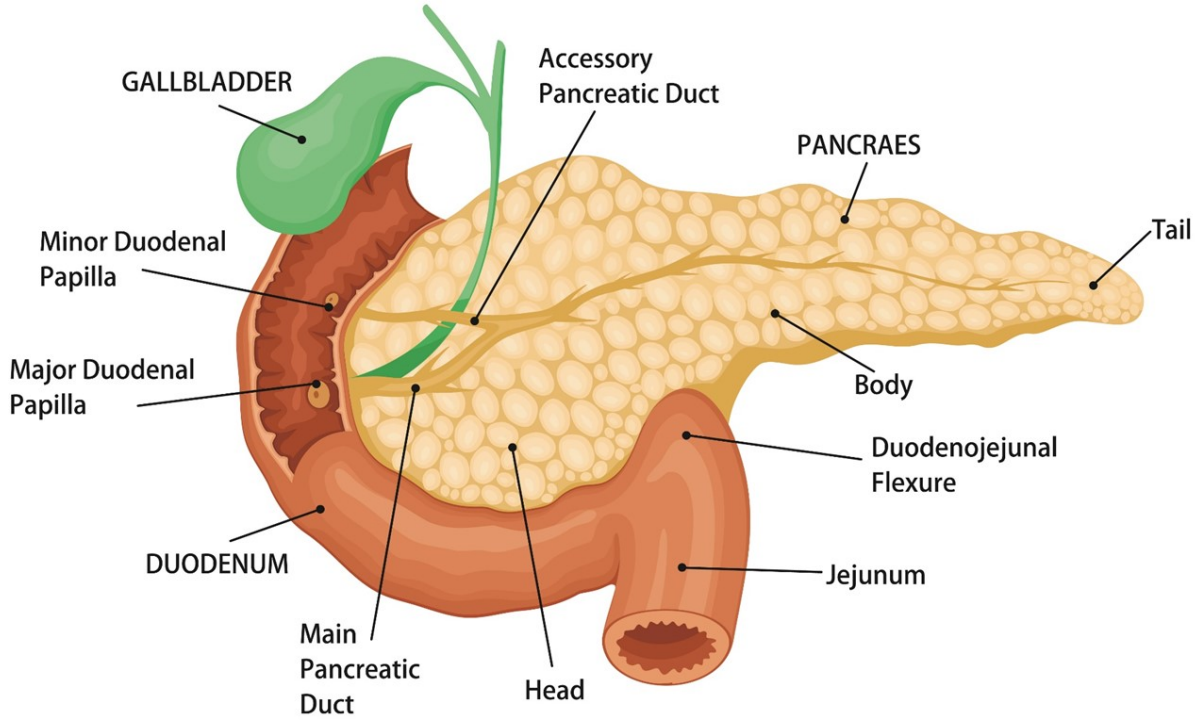


Figure 1.1: Anatomical depiction of the pancreas and major interaction partner. While different cells in the pancreas can develop into tumours, most pancreatic malignancies begin in the pancreatic ducts in the head of the pancreas. Purchased from creative market[7].

The less common neuroendocrine tumours arise from the hormone-producing endocrine system and have different clinical characteristics as well as a more favourable prognosis[8]. Pancreatic neuroendocrine tumours (PanNETs) are grouped into functional and non-functional types depending on their ability to maintain hormone production after tumour formation. Functioning PanNETs are often diagnosed earlier because they typically produce and secrete hormones (e.g. insulin or gastrin) at a higher rate, which can cause serious symptoms[9].

1.1.1.2 Diagnosis and classification

One of the major reasons for the short OS in pancreatic cancer is its late and often unspecific symptoms. They are related to the tumour location within the pancreas and most commonly include abdominal pain, which may worsen at night; weight loss; asthenia; anorexia; and jaundice[10,11]. Additionally, diabetes is present in at least 50% of patients[12]. With the unspecific nature of the symptoms, the diagnosis of pancreatic cancer is mostly based on medical imaging such as computed tomography (CT) and endoscopic ultrasound (EUS). Endoscopic ultrasound can be combined with fine-needle aspiration to extract a small tumour biopsy that helps with tumour classification and the assessment of its respectability[13]. In many pancreatic cancer patients, the level of Carbohydrate antigen 19-9 (CA19-9) is also checked. The cell surface molecule can often be found in cancer patients using elevated serum concentrations[14]. While CA 19-9 levels have been criticized for possessing a low sensitivity that leads to false-

positive and false-negative prognoses, they can be used in concert with diagnoses made with medical imaging, can predict prognosis in respectable tumours, and have been shown to correlate with tumour size, stage, and burden[15-17]. The level of CA 19-9 was also shown to be a reliable biomarker for the detection of pancreatic cancer reoccurence[18]. In recent years, liquid biopsy has been proposed as an additional easy and non-invasive method for prognosis, classification, and disease monitoring.

After a pancreatic cancer diagnosis, tumours and patients' overall fitness are graded based on several different classification systems. The staging system most often used for exocrine pancreatic cancer is the American Joint Committee on Cancer (AJCC) TNM system, which is based on three key pieces of information: extent of the tumour (T), spread to the nearby lymph nodes (N), and spread (metastasizes) to distant sites (M; Table 1.1).

Table 1.1: Definition of the 8th edition of TNM staging system of PDAC by AJCC[19].

TNM system		AJCC stages			
T1	Maximum tumour diameter ≤2 cm		**T**	**N**	**M**
T2	Maximum tumour diameter >2, ≤4 cm	**IA**	T1	N0	M0
T3	Maximum tumour diameter >4 cm	**IB**	T2	N0	M0
T4	Tumour involves the celiac axis, common hepatic artery or the superior mesenteric artery	**IIA**	T3	N0	M0
N0	No regional lymph node metastasis	**IIB**	T1-T3	N1	M0
N1	Metastasis in 1–3 regional lymph nodes	**III**	T4 (any T)	Any N (N2)	M0
N2	Metastasis in ≥ 4 regional lymph nodes	**IV**	any T	Any N	M1
M0	No distant metastasis				
M1	Distant metastasis				

Additionally, tumour grade, which is based on how much the altered tissue resembles its tissue of origin, and the Karnofsky score, which classifies the patients' overall fitness, have applied for PDAC evaluation. The combination of the classification results is important for risk stratification, outcome prediction, and choice of treatment modality.

1.1.2 Epidemiology

1.1.2.1 Incidence and mortality

While not exceedingly common, pancreatic cancer is one of the deadliest types. It is ranked as the 14th most common cancer and the 7th highest cause of cancer death worldwide. The world health organisation's official index counted 458,918 new diagnoses and 432,2242 deaths from pancreatic cancer in 2018[20]. The incidence of pancreatic cancer differs widely among

areas and populations (Figure 1.2). The age-standardised rate (ASR) was highest in Europe (7.7 per 100,000 people) and North America (7.6 per 100,000 people), and the lowest in Africa (2.2 per 100,000 people) and Western Asia (4.9 per 100,000 people)[21]. The general trend of higher incidence rates in developed countries than developing countries is thought to be due to different exposure to risk factors and developed countries' improved diagnostics tools[22].

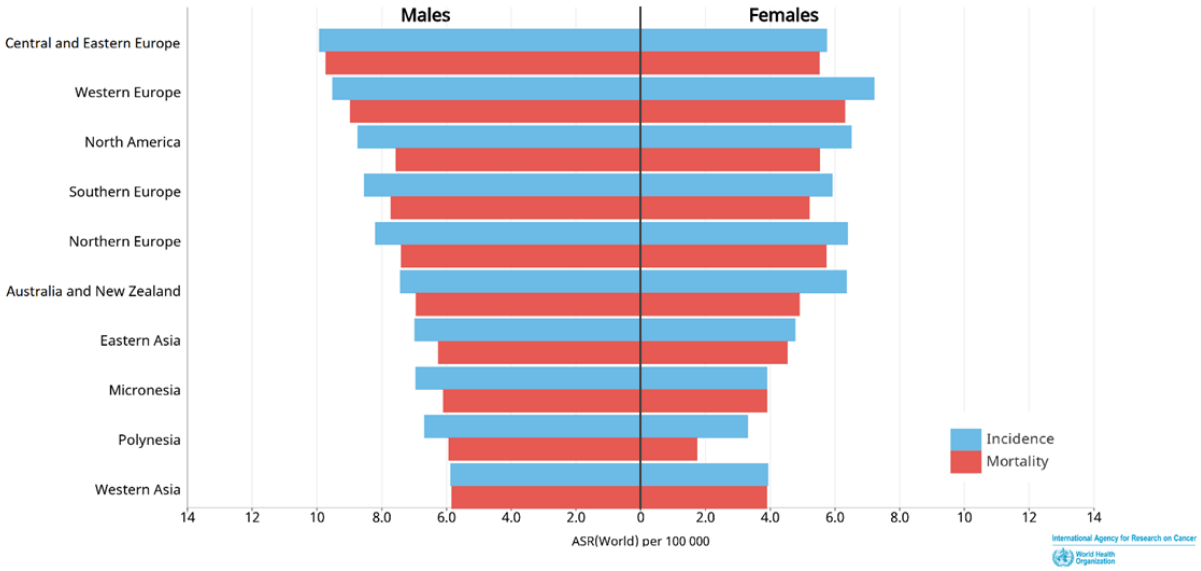


Figure 1.2: Estimated age-standardised incidence and mortality rates for pancreatic cancer in 2018 worldwide. With mortality almost equalling incidence, pancreatic cancer has one of the worst prognoses of all cancer types. It is more prevalent in developed countries and occurs more often in men than in women. Age-standardized rates are shown across all ages. Modified from Global Cancer Observatory (http://gco.iarc.fr/).

Pancreatic cancer is slightly more common in men, and the incidence rate for both genders increases with age. People are rarely diagnosed before 55 years of age, and the highest incidence is reported in people over 70 years[23].

The close parallel between incidence and mortality shows how deadly pancreatic cancer is – a fact that is reinforced by the five-year-survival rate of only 10%[24]. Due to late and unspecific symptoms, over 80% of patients are diagnosed at advanced stages, making curative intended pancreatectomy impossible[25]. However, even patients with resectable tumours relapse in the majority of cases (80%), leading to a five-year-survival rate of around 30%[26]. Even though improvements in diagnosis and treatment have slightly increased the five-year survival rate in recent years, both the number of cases and the number of deaths from pancreatic cancer is projected to increase significantly over the next decades[20]. This translates to a predicted rise pancreatic cancer from the fourth- to the second-most common cause of cancer-related deaths in the United States by 2030[27,28]. Reasons for this might be increased age, changes in lifestyle, and, subsequently, increased exposure to pancreatic cancer risk factors and the lack of potent therapeutics responsible for lower cancer death rates in other tumour entities[29].

1.1.2.2 Aetiology and risk factors

The poor prognosis and delayed diagnosis of pancreatic cancer make prevention particularly important. A long list of risk factors have been identified; these can be further divided into modifiable and non-modifiable risk factors. The most important environmental modifiable risk factor is smoking, as duration of smoking and the number of cigarettes smoked daily increases the risk of a pancreatic cancer diagnosis[30-32]. Excessive alcohol consumption itself increases pancreatic cancer risk[33] and is also a main cause of chronic pancreatitis, which is another known risk factor for pancreatic cancer[34]. Approximately 5% of patients with chronic pancreatitis will develop pancreatic cancer during their lifetimes[35]. The final two major modifiable risk factors for pancreatic cancer are obesity[36,37] and exposure to carcinogens such as cadmium[38,39] and nickel[40].

Risk factors that are not modifiable include gender, age, ethnicity, diabetes mellitus type 1, family history of pancreatic cancer, and non-O blood group[21]. While the increased incidence in men might be explained by differences in lifestyle such as heavy smoking and high alcohol consumption, there may yet be undiscovered biological factors responsible for the disparity. Similarly, differences in pancreatic cancer incidences among different ethnicities can be explained by a combination of risk factor exposure and genetic factors[41-43]. Diabetes mellitus is both a risk factor and an early symptom of pancreatic cancer. Several large meta-analyses have shown a 1.8–2 times increased risk of developing pancreatic cancer in diabetes patients[44,45]. Approximately 5–10% of all pancreatic cancer cases have a family history of the disease, meaning that at least two first-degree relatives have previously been diagnosed[46]. The risk for pancreatic cancer increases with the amount of family members affected[47]. Finally, the blood groups A, AB, and B have all shown to have a 1.3–1.7 times increased risk of developing pancreatic cancer[48].

1.1.3 Pathophysiology of PDAC

As is common for many cancer types, pancreatic cancer evolves through non-invasive precursor lesions. The three best-characterised of those lesions are pancreatic intraepithelial neoplasia (PanIN), intraductal papillary mucinous neoplasms (IPMN), and mucinous cystic neoplasms (MCN). Pancreatic intraepithelial neoplasia is the most common precursor of PDAC and is commonly classified into low-grade (formerly grade 1a/b and 2) and high-grade PanIN (formerly grade 3)[49]. In the classical model of PDAC, development low-grade PanINs are considered the disease-initiating event. Though these lesions harbour a high frequency of *KRAS* mutations (> 90%), an increased mutations burden, enhanced DNA damage response, short telomeres, and an active senescence program (as compared to normal tissue)[50,51], they are generally considered benign because they can be found in over 70% of older people with no evidence of pancreatic cancer[52]. Therefore, *KRAS* mutations alone appear insufficient to

overt full-blown carcinogenesis. Inactivation of the *CDKN2A* locus was demonstrated in the form of epigenetic silencing, mutation, or CNA in 30% of low-grade PanIN but 80% of high-grade PanIN. The gradual increase of *CDKN2A* alteration suggests that inactivation of the tumour suppressor is required for high-grade PanIN formation[53]. The other two most frequently altered gene aberrations of PDAC, *TP53* and *SMAD4*, are only rarely found in low-grade PanIN, but they are often detected in high-grade lesions[54,55]. Since high-grade PanINs are the equivalent of carcinoma *in situ*, *TP53* and *SMAD4* inactivation appears to be a late step in PDAC development (Figure 1.3).

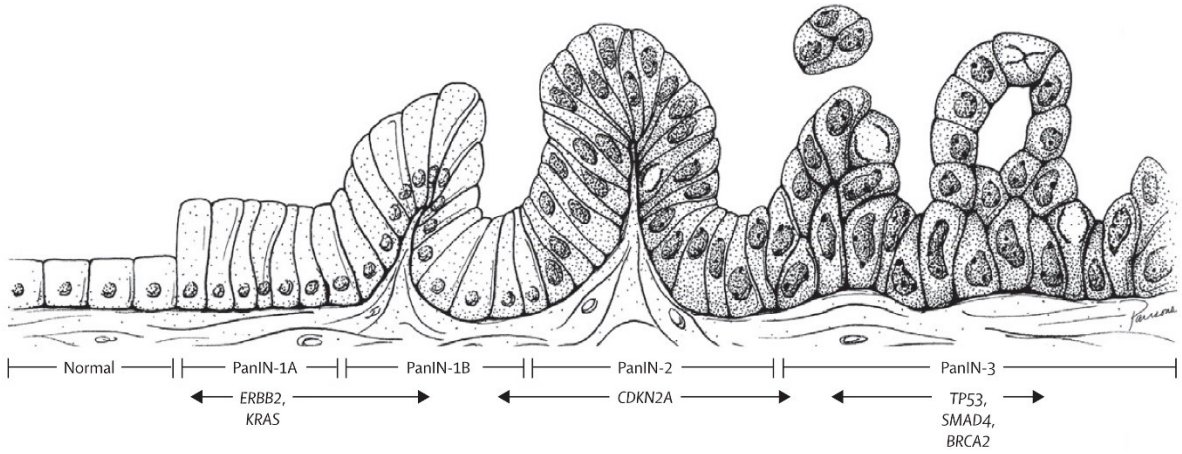


Figure 1.3: Classical PDAC progression model including genetic alterations. In the evolution from normal pancreatic tissue over low-grade (PanIN-1A/B and PanIN-2) to high-grade PanINs (PanIN-3), the early *KRAS* mutations are an early step towards invasiveness. Nevertheless, additional inactivation of one or more tumour suppressors is necessary for the tissue to reach a cancer state. Examples for typical gene alterations are shown for the different lesion in the PDAC progression model. Reprinted from Hruban et al.[56].

Comparative genomic analyses of low-grade PanINs, primary PDAC, and matched metastasis have offered a time frame of nearly 20 years from the first initiating *KRAS* event to a fully metastatic disease[57,58]. Close relationship between distant metastasis and initial tumour side imply metastasis formation to be a ‘late’ event in tumour development[58,59]. Combined with the considerable intra-tumour heterogeneity of invasive PDAC, this has led to the hypothesis that a final catastrophic event like chromothripsis or chromoplexy transforms benign pancreatic lesions into locally invasive tumours with metastasis-forming abilities[60]. Even accounting for the large increase in knowledge of the genetic background of PDAC, the final model of PDAC progression and metastasis formation remains a matter of debate. Large-scale, single-cell investigations appear promising methods for elucidating the remaining questions[61,62].

1.1.4 Pancreatic cancer microenvironment

One of the main characteristics of pancreatic cancer is its dense microenvironment, which can comprise up to 80% of the tumour mass[63]. The stroma consists of a complex mixture of proliferating myofibroblasts (pancreatic stellar cells, PSCs); multiple types of inflammatory cells

including macrophages, mast cells, lymphocytes and plasma cells; and different extracellular matrix (ECM) components[3]. The cancer microenvironment interacts with the tumour cells by producing growth factors and immune suppressants as well as by promoting metastasis formation. Its two main features are dense desmoplasia and strong immunosuppression[64]. Desmoplasia describes an extensive fibrosis at the primary tumour site, which is caused by overexpression of ECM proteins and the transformation of fibroblastic-type cells to a myofibroblastic phenotype[65]. It increases tumour progression and resistance to chemotherapy and is therefore associated with a poor prognosis[66]. In combination with insufficient vasculature, the dense stroma create a hypoxic environment in addition to secreting growth factors. This contributes to pancreatic cancer aggressiveness, including metabolic reprogramming, inhibition of apoptosis, sustained proliferation, treatment resistance, invasion, and metastasis[67]. The second major impact of the tumour microenvironment is the restriction of immune surveillance and creation of an inflammatory program that supports tumour formation trough crosstalk between tumour and immune cells[68]. Blockage of the T-cell mediating anti-tumour immunity is activated during early stages of tumour formation through recruitment of regulatory T-cells and myeloid-derived suppressor cells[69]. Additionally, pancreatic cancer has shown to decrease its MHC1 expression, which in turn prevents recognition by killer T-cells (CD8+)[70]. Moreover, immunosuppressive macrophages are recruited into the microenvironment where they secrete immunosuppressive cytokines, chemokines, and enzymes such as TGFβ, IL-10, CCL17, and CCL22[71,72]. These properties of the tumour microenvironment are believed to be a major cause for the frequent failure of anticancer immunotherapy in pancreatic cancer. In addition to desmoplasia and immunosuppression, the tumour microenvironment also increases tumour resilience by interacting with pancreatic cancer stem cells, favouring their stemness maintenance, which includes self-renewal, tumorigenic, and metastatic potential[73-75]. Finally, the pancreatic cancer microenvironment facilitates metastasis formation via multiple signalling pathways[66]. Immunosuppressive macrophages can secrete cytokines that activate tumour-associated lymphangiogenesis, which directly correlates with pancreatic cancer lymphatic metastasis and OS[76,77]. Immune cells residing in the microenvironment have shown to promote epithelial-mesenchymal transition (EMT) in pancreatic cells[78-80], and exosomes derived from the pancreatic cancer microenvironment can induce a pre-metastatic niche formation in the liver and lungs[81,82]. In summary, pancreatic neoplasm consists not only of invasive cancer cells but also includes a dense and complex microenvironment that facilitates tumour progression, immune suppression, cancer stem cell maintenance, and metastasis formation. Therefore, the PDAC tumour microenvironment has increasingly been recognized as a promising target for cancer precision medicine.

1.1.5 Current treatment options for PDAC

1.1.5.1 Classical treatment strategies

For pancreatic cancer patients with resectable tumours restricted to the pancreas and no involvement of adjacent organs or vessels (stage I and II), the treatment of choice is surgical removal. The recurrence rates after resection of are high, up to 80%, and have warranted the use of additional systemic chemotherapy either before (neo-adjuvant) or after surgery (adjuvant)[83]. In 2013, the CONKO-001 study compared adjuvant treatment with gemcitabine with observation. Adjuvant treatment with gemcitabine resulted in a significant improvement of DFS, proving the benefit of adjuvant chemotherapy[84]. Subsequent trials have established superiority of combinatorial regimens such as modified FOLFIRINOX (fluorouracil, leucovorin, irinotecan, and oxaliplatin; mFOLFIRINOX) over gemcitabine alone, which is accompanied with a higher incidence of toxic effects[85]. However, gemcitabine remains a recommended option for patients with comorbidities. The role of neo-adjuvant chemotherapy in PDAC remains a topic of discussion. It is most often used to treat borderline resectable PDAC in the hope of reducing tumour margins and enabling surgical treatment of the disease. Other aims are an increased R0 resection rate and elimination of micrometastasis hopefully leading to lower relapse rates and increased DFS[86]. A combination of data from several single-arm studies and a recent randomized phase III trial (PREOPANC[87]) indicated benefits of neo-adjuvant treatment over upfront surgery with adjuvant treatment alone, especially in borderline resectable PDAC[88].

For locally advanced and metastatic disease, palliative chemotherapy is the standard of care. Additionally, pain management and symptom control, such as management of jaundice via biliary drainage, are the main concerns in advanced pancreatic cancer. For decades, the only chemotherapeutics available were fluorouracil (5-FU) and gemcitabine with progression-free survival (PFS) and OS rates of 4–6 months[89]. The first real improvement to chemotherapy regimens came in 2011 with the introduction of FOLFIRINOX, which showed a superior PFS and median OS as compared to gemcitabine monotherapy[90]. Unfortunately, the improved outcome is accompanied by increased toxicity and is therefore limited to patients with good performance status. A second, less toxic, standard chemotherapy for advanced PDAC is the combination of gemcitabine with nab-paclitaxel. In 2013, the phase III MPACT trial showed a significant increase of PFS and OS rates compared to gemcitabine alone[91].

1.1.5.2 Targeted therapy and immune therapy

As with other solid tumours, recent research in PDAC has focused on targeted therapy approaches using small molecules. However, only limited success has been achieved using targeted therapies in PDAC patients. The Know Your Tumour initiative did demonstrate the

feasibility of precision medicine as part of the treatment of PDAC; however, by matching targeted therapies to patients with actionable lesions, it only identified 35 out of 1,020 patients (3%) for targeted therapy, resulting in only a modest improvement to DFS[92]. The need for additional actionable lesions and therapies that are more potent is currently being answered in a wide range of clinical studies targeting signalling pathways important in PDAC progression; regulators of the DNA repair mechanism; and key players in cellular metabolism, tumour microenvironment, and pancreatic cancer stem cells. With *KRAS* mutated in over 90% of patients, the adjacent signalling cascade was a clear starting point. In 2019, a novel inhibitor of *KRAS*^{G12C} (AMG510) showed to generate disease control in advanced-stage *KRAS*^{G12C}-mutant solid tumours in an early phase trial[93]. However, these mutations accounted for only ~1% of all KRAS mutations observed in PDAC. Thus, to this day, a meaningful inhibition of KRAS itself is awaited. In 2007, an inhibitor of one of its upstream receptors – the epidermal growth factor receptor (EGFR) – was shown to significantly increase PFS and OS from gemcitabine monotherapy when used in combination with gemcitabine[94]. Even though a wide range of other agents have been or are currently being tested, the tyrosine kinase inhibitor erlotinib is the only targeted therapeutic that has been approved by the federal drug administration (FDA). Increased knowledge of the genetic background of PDAC has led to many potentially targetable pathways that are altered, like Notch, Hedgehog, mTOR and JAK/STAT. So far, however, all tested inhibitors have failed in clinical trials[95]. One explanation for this disappointing track record is the physical barrier of PDAC's dense stroma with its few blood vessels and strong immunosuppression. While one recent clinical study using hyaluronic acid (HA) degrading enzyme PEGPH20 in combination with nab-paclitaxel/gemcitabine did show some promising results in high HA patients,[96] therapy options for PDAC are still limited and the outcome remains bleak.

In contrast to other solid tumours, immunotherapy has shown disappointing results in PDAC so far. This is possibly due to the immune suppressive tumour microenvironment (with stromal infiltration of myeloid-derived suppressor cells and tumour-associated macrophages) and low mutational burden as compared with other cancer entities such as melanoma[97]. One area of ongoing research in PDAC immunotherapy is immune checkpoint inhibitors, which showed disappointing results in early trials but which are now evaluated in combination with immunomodulatory agents like cancer vaccines[98]. Furthermore, adoptive T-cell transfer with T cells engineered to transiently express a mesothelin-targeting CAR has shown limited response in a phase 1 trial and is one of several chimeric antigen receptor (CAR) T-cell therapies currently being tested in advanced PDAC[98,99]. Finally, tumour-necrosis factor (TNF) receptor CD40, which is expressed on different immune cells including antigen-presenting cells (APCs) and some tumour cells, has been targeted with CD40-specific agonists to enable APCs to increase T-cell activation level[98]. Despite some encouraging preliminary results, to date no

immunotherapy regime has been approved for PDAC treatment, and more in-depth analysis of PDAC biology might be necessary to overcome the immune barrier of this deadly disease.

1.2 Precision medicine in PDAC

In recent decades, tumour heterogeneity has been increasingly recognized as both obstacle and chance for cancer treatment. Advances in next-generation sequencing (NGS) technologies has led to a tremendous increase in the knowledge of the molecular background of almost all cancer types. Unfortunately, patients with pancreatic cancer have not yet benefitted from recent improvements in precision medicine, as seen in other malignancies.

1.2.1 Genetic background of PDAC

Several large NGS studies have provided a much clearer picture of PDAC biology. Analysis of CNA, structural variants (SV), and SNV have identified a relatively small set of four core-mutated genes (*KRAS, TP53, CDKN2A,* and *SMAD4*) and a long tail of rather infrequently altered genes (≤10% mutation frequency). Among the most common ones are *KDM6A, ARID1A, TGFBR2, KMT2C, RNF43* and *ATM*[100-102]. A number of these genes affect the same pathways and biological processes, including NOTCH, Hedgehog, β-catenin, axon guidance, chromatin remodelling, and DNA repair pathways. This suggests that the majority of these mutations may function through certain core mechanisms, which may in turn offer points for therapeutic intervention[102,103]. Approximately 10% of all cases, mainly in patients with a family history of PDAC, uncover germline mutations affecting the DNA damage repair machinery such as *BRCA1, BRCA2, PALB2, ATM, CDKN2A, APC, MLH1, MSH2, MSH6, PMS2, PRSS1,* and *STK11*[47,104]. Additionally, patients with familial pancreatic cancer typically have more precancerous lesions than those with sporadic pancreatic cancers[105] and are linked with a number of specific syndromes associated with the same germline alterations[106].

1.2.2 PDAC subtypes

As a reaction to the apparent inter-individual heterogeneity of PDAC, several attempts have been made to classify patients into groups based on their genetic and transcriptional profiles. Two main classification approaches have been postulated in the past. First, using whole-genome sequencing of 100 PDACs, Waddell and colleagues classified PDAC into four subtypes based on structural rearrangements: (1) ‘stable’, with less than 50 structural variations; (2) ‘locally rearranged’, defined by the presence of a significant focal event on one or two chromosomes including amplifications in known oncogenes such as *KRAS, SOX9* and *GAT6*; (3) ‘scattered’ for tumours with non-random chromosomal damage and less than 200 structural events; and (4) ‘unstable’ for patients with more than 200 structural variants[101]. The unstable genotype was linked to alterations in DNA damage repair genes such as *BRCA1* and *BRCA2* and was subsequently shown to be especially sensitive to platinum-based therapy[101].

Simultaneously, Collisson et al[107]., Moffit et al.[108], and Bailey at al.[102] proposed 2–4 clinically relevant PDAC subgroups based on transcriptional differences with overlapping definitions.

The subgroups have significant prognostic power and can predict disease outcomes. For example, the quasi-mesenchymal (Collisson et al.), basal-like (Moffit et al.), and squamous (Bailey et al.) subtypes are comparable across all three classification systems and have significantly shorter OSes than the rest[102,107,108]. A recent meta-analysis even postulated possible predictive implications for the mRNA based subtypes[109] .Based on 17 patient-derived mouse xenografts, they found significant, subtype-specific differences in sensitivity to gemcitabine and 5-FU[110].

1.2.3 Prognostic and predictive factors genetic in pancreatic cancer

A multitude of clinical studies have discussed the prognostic and predictive utility of aberrations in individual genes. However, few of these studies have been independently validated, and virtually no genetic marker has been adopted in clinical practice. Due to their high prevalence, the four most frequent genes in PDAC (*KRAS, TP53, CDKN2A and SMAD4*) have been at the centre of the search for prognostic biomarkers. Several groups have investigated whether the presence of *KRAS* mutations harbours prognostic information. Keeping important differences such as detection methods and cohort size in mind, most studies concluded that *KRAS* mutations are associated with an adverse effect on survival[111] – particularly when leading to amino acid alterations at codon 12 (G12D, G12R or a combination of two mutants)[112]. Both *TP53* and *CDKN2A* have been repeatedly studied to determine their impacts on disease outcomes. While several studies have demonstrated a link between their mutational status and poor clinical outcomes, no final conclusions have been made as to their applicability as prognostic biomarkers[113]. The most promising prognostic markers with potential clinical utility are loss of *SMAD4* expression and overexpression of *S100A2*. Both are suitable for detection with immunohistochemistry-based assays and are associated with a poor prognosis and a pattern of more disseminated disease[114]. Clinical effects of the reduction of *SMAD4* expression in patients with resectable PDAC have been extensively studied using either immunohistochemistry or sequencing. In the majority of those studies, loss-of-function mutations, deletions, and low protein levels have been significantly related to poor prognoses[115]. Its prognostic power in advanced PDAC has, however, recently been questioned[116]. Additionally, patients with *SMAD4* loss are commonly associated with distant disease progression, whereas patients with wild-type *SMAD4* harbour a local tumour pattern of progression[117]. High *S100A2* expression levels have proven an independent predictor of survival in a multivariate model with clinical variables[118], even predicting adjuvant therapy benefits in PDAC[119]. Recently, a prognostic risk score based on expression levels of *S100A2* and *S100A4* stratified PDAC patients into three distinct prognostic groups[120].

The search for predictive biomarkers that guide treatment approaches in PDAC has produced few robust results. Even genetic marker approaches that have shown an effect in other tumour

entities could not be implemented in clinical practise for PDAC treatment. One exception is the approval of pembrolizumab for first-line treatment of any solid tumour with microsatellite instability (MSI). Occurring in 1–2% of resectable PDAC, MSI is the result of a functional loss of several DNA mismatch repair genes (MSH1, PMS2, MLH1 and MSH6)[121]. Other predictive biomarkers applicable in multiple cancer types (e.g., *ERBB2* amplification, *BRAF,* or *BRCA1*/*2* mutations) occur in PDAC at low prevalence even with promising preclinical data and have yet to be proven beneficial in a clinical setting[101,114]. The only potential biomarker for standard chemotherapy treatment in PDAC, hNET1 for gemcitabine responsiveness, has also been taken out of consideration after a phase III trial showed non-significant benefits for gemcitabine adjuvant therapy in *hNET1* high expression patients[122]. Erlotinib, the only approved targeted therapeutic for advanced PDAC, would particularly benefit from biomarker-driven patient subgrouping. This is due to its small survival benefit of only two weeks, which makes its use in clinical practice unlikely. Well-established predictors in other solid tumours – such as *EGFR* in erlotinib sensitive lung cancer, which could have been applicable for PDAC – failed to show any predictive power[123]. Cell lines with the classical Collisson subtype, which is enriched for *GATA6* mutations, exhibited increased sensitivity toward erlotinib treatment[107]. Therefore, *GATA6* expression has been proposed as a predictive biomarker for erlotinib response. Thus far, this has not been validated in PDAC patients.

1.3 Important signalling pathways affecting erlotinib treatment in PDAC

Understanding the major signalling pathways responsible for different aspects of tumour progression and treatment evasion is key in any attempt to improve available treatment regimes, overcome resistance, and find new targets. Blocking the epidermal growth factor receptor (EGFR) with the small molecule erlotinib affects a complex net of downstream signalling pathways that can interact with each other to change cancer cell behaviour.

1.3.1 EGFR signalling

Human EGFR type 1 (HER1/EGFR) is a transmembrane glycoprotein belonging to the HER family of receptor tyrosine kinases (RTK; also called ErbB receptors). After activation by binding its ligands, epidermal growth factors (EGFs) and transforming growth factor alpha (TGFα) EGFR forms activated homodimers, which stimulates the intrinsic intracellular protein-tyrosine kinase activity. As a result, intracellular autophosphorylation of several tyrosine residues occurs. Binding of erlotinib to the ATP binding site of the receptor inhibits the phosphorylation of the tyrosine residue. Over the docking of several proteins, such as GRB2 and SOS, to the phosphorylated residues, the signal is then relayed toward intracellular signalling cascades. The transforming growth factor alpha activates two major signalling pathways (Figure 1.4). First is the mitogen-activated protein kinase (MAPK) pathway, which affects the downstream target MAPK and directs gene transcription, cell cycle progression, and cell proliferation. Second is the phosphatidylinositol-3 kinase (PI3K)/ protein kinase B (PI3K/Akt) pathway, which affects the downstream activation of the mammalian target of rapamycin (mTOR) and nuclear factor κB (NF-κB) and generates anti-apoptotic and pro-survival signals[124]. The EGFR has shown to be overexpressed in pancreatic cancer, a phenomenon linked to structural or numerical alterations of chromosome 7[125]. Activating EGFR mutations are found in less than 2% of PDAC patients[102]. Increased EGFR expression has been associated with shorter OS in some but not all PDAC studies[126], and EGFR inhibitors have been tested for the treatment of advanced PDAC in several clinical studies. The highly dynamic nature of EGFR and subsequent MAPK signalling makes targeted treatment approaches complicated. Therefore, of those targeted therapeutics, only erlotinib has shown to exhibit a small improvement in OS[94,127].

1.3.2 MAPK Pathway

The MAPK pathway is one of the central signalling pathways in cell survival and growth. The classical signal transduction begins with activation by a receptor tyrosine kinase, such as EGFR, which then promotes the exchange of GDP to GTP in the intracellular GTPase Ras (rat sarcoma viral oncogene homolog, most notably HRas or KRas). Ras-GTP activates MAP3K (e.g., rapidly accelerated fibrosarcoma Raf), which phosphorylates MAP2K (e.g., MAPK

extracellular signal regulated kinases, MEK1/2), which in turn phosphorylates MAPK (e.g., extracellular signal-regulated kinase ERK1/2). Phosphorylated MAPK has over one hundred downstream cytoplasmic and nuclear substrates[128]. It can phosphorylate several different cytoplasmic effectors, thereby activating their signalling pathway, such as PI3K/Akt, or crossing into the nucleus to directly regulate transcription factors (e.g., c-myc and cAMP response element-binding protein, CREB; Figure 1.4). In PDAC, activating *KRAS* mutations have been found in more than 90% of patients and PanINs[50,102]. However, in the absence of additional deleterious events in strong tumour suppressors, activating *KRAS* mutations have shown to induce senescence[129].

One of the main signalling cassettes of the MAPK pathway is the c-Jun N-terminal kinase (JNK) pathway. Activated by environmental stresses (ionizing radiation, heat, oxidative stress, and DNA damage) and inflammatory cytokines as well as growth factors, JNK signalling can be initiated via RTKs as well as through G-protein-coupled transmembrane receptors (GPCRs). Signalling subsequently goes through KRAS, different MAP3Ks, MAP2K4/MAP2K7, and ends with activated JNK. Like the other MAPKs, JNKs translocate into the nucleus where they tend to physically associate with their target transcription factors such as c-Jun, ATF, Elk1, SMAD4, and p53 to activate them[130]. While the exact mechanism trough which JNK signalling influences PDAC progression is not yet known, it has shown to effect chemoresistance and cancer cell invasiveness in PDAC cells[131,132].

1.3.3 PI3K/Akt Pathway

One of the main signalling partners of the MAPK pathway is the PI3K/Akt pathway. Mmuch like Ras, Akt is a strong proto-oncogene that activates cell growth and proliferation via a multitude of downstream signalling partners. The PI3K/Akt pathway can be activated by both RTKs and GPCRs. As a result, PI3K activates and creates the membrane-bound phosphatidylinositol (3,4,5) trisphosphates (PIP_3). These serve as plasma membrane docking sites for proteins harbouring pleckstrin-homology (PH) domains, including Akt and its upstream activator PDK1. After phosphorylation and activation by PDK1 (3-phosphoinositide-dependent protein kinase-1), Akt has a number of downstream effects such as activating CREB, inhibiting p27, localizing the forkhead-box-protein (FoxO) in the cytoplasm, activating the NFκB pathway, and activating mTOR[133]. To regulate the strong effect on cell survival and cell-cycle progression Akt controls, several regulators are established. For example, the tumour suppressor phosphatase and tensin homolog (PTEN) inhibit Akt activity by dephosphorylating PIP_3 (Figure 1.4). Separate studies reported that around 50% of pancreatic cancers show increased PI3K signalling activation, as assessed with the phosphorylation of Akt. This has been associated with an undifferentiated grade and correlated with poor prognoses[134]. Mutations in the PI3K/Akt pathway molecules are rare in PDAC. For instance, PIK3CA was found mutated in only about

4% of patients[101], meaning that other mechanisms must be responsible for the consistent activation. Common aberrations of important players in the PIK/Akt pathway include activating mutations of *KRAS* (which has a gene product known to induce PI3K/Akt/mTOR) [102], Akt2 amplification in 10–32% of PDAC patients[135] and functional inactivation of PTEN via point mutations in its gene, and deletions. Additionally, upregulations of growth factor pathways and activation of GPCR signalling (through the expression of their specific ligands such as Shh or gastrointestinal peptide hormones) have all been seen in PDAC and might explain the pathway activation[136,137].

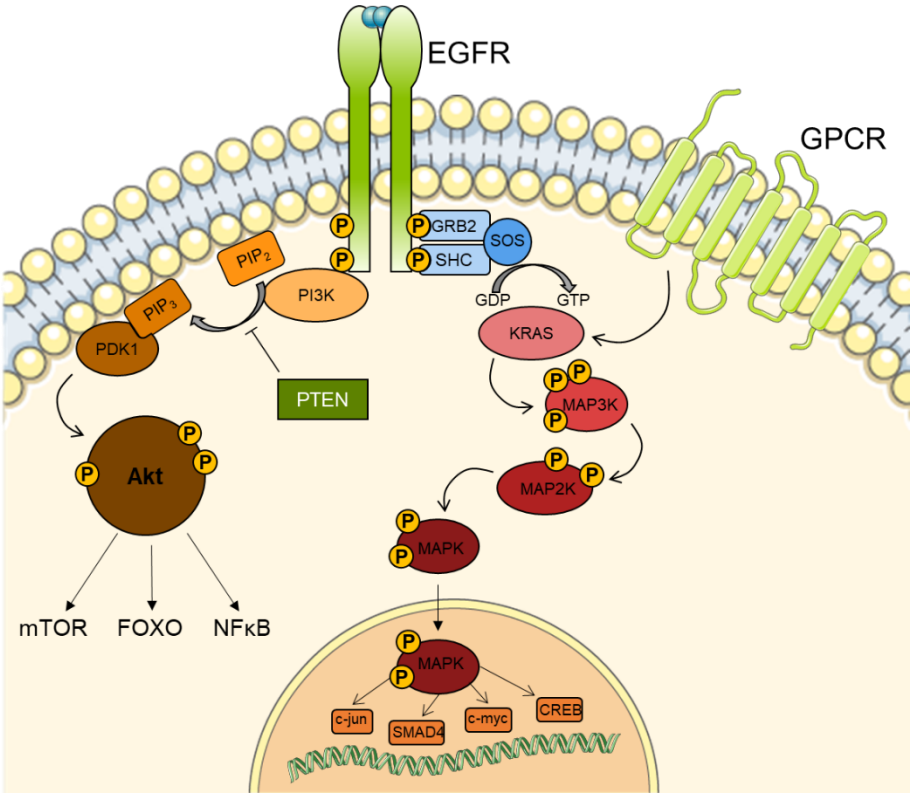


Figure 1.4: Overview of the major signalling pathways downstream of EGFR. After binding its ligand the receptor autophosphorylates and subsequently activates the downstream effectors in the MAPK and PI3K/Akt pathways.

1.3.4 TFGβ signalling pathway

The transforming growth factor beta (TFGβ) signalling pathway is a complex network with widely varying roles in cellular homeostasis. Signal relay begins with the binding of TGFβ superfamily ligands (e.g., TFGβ, bone morphogenetic proteins [BMP], and activin) to a type II receptor (e.g., TGFβR2, BMPR2, or ACVR2A/B), which is followed by dimerization and recruitment of type I receptor dimers. The type I receptor then phosphorylates receptor-regulated SMADs (R-SMADs), which can then bind the coSMAD SMAD4. R-SMAD/coSMAD complexes accumulate in the nucleus where they act as transcription factors. Additionally, TGFβR can translate signals through a SMAD4-independent, non-canonical signalling cascade, thereby activating the MAPK pathway and the PI3K/Akt pathway[138] (Figure 1.5). The

effects of TGFβ signalling range from tumour suppression to proto-oncogenic activities during cancer progression. TGFβ exhibits potent growth inhibitory effects in early stages of pancreatic cancer by promoting apoptosis, differentiation, and G1 arrest. In contrast, during advanced stage of carcinogenesis, TGFβ promotes invasion and metastasis of pancreatic cancer as well as stromal growth factor production and angiogenesis while suppressing an anti-tumour immune response[139]. The TGFβ pathway belongs to a set of core signalling pathways mutated in at least one gene in virtually every of PDAC patient[140]. These mutations can be found to a small extent in the pathway receptor genes, but they are predominantly the result of mutations or deletions of the coSMAD SMAD4 expressing gene, which affects up to 90% of PDAC patients[102,141].

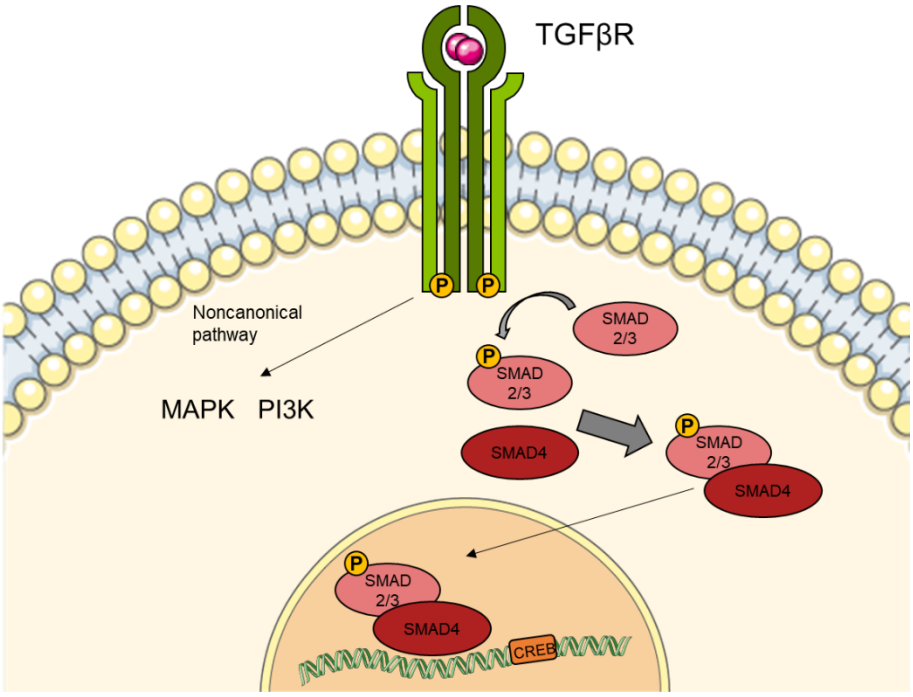


Figure 1.5: Overview of TGFβ signalling pathway. Signal relay through the canonical signalling pathway is activated when the ligand binding leads to formation of the receptor tetramer and subsequent intracellular phosphorylation. The rSMADs (SMAD2/3) are in turn phosphorylated and bind to the coSMAD SMAD4. The pair translocates into the nucleus and directly interacts with DNA of their downstream effectors to facilitate their transcription.

1.4 Aim

PDAC is one of the deadliest types of cancer with a mortality rate nearly as high as its case rate. Late detection and resistance to both classical and targeted therapy approaches are major reasons for the dismal prognosis. In recent years, several large sequencing studies have increased our knowledge of its molecular biological background. Despite this, the only approved targeted therapeutic to date is the tyrosine kinase inhibitor erlotinib. Its positive effect on OS is small, and, without specific biomarkers to identify patients who will benefit more extensively, its clinical use is limited. With all but the core PDAC gene (*KRAS*, *TP53*, *CDKN2A*, *SMAD4*) mutated in less than 5% of patients, effects of low frequency alterations are easily lost. Therefore, the small size of and strong heterogeneity within sample cohorts used in clinical trials and previous large-scale sequencing studies would have hindered earlier attempts to find predictive and prognostic biomarkers for PDAC. To address this knowledge gap, I performed an in-depth molecular characterization of 293 R0-resected PDAC patients treated within a multicenter phase III trial. In this trial, adjuvant chemotherapy of gemcitabine – with or without erlotinib – was compared. While the addition of erlotinib failed to improve outcomes in the entire study population, its efficacy in some patients warrants predictors of responsiveness. The goal of my sequencing approach was to attain a detailed genetic analysis of R0-resected PDAC, which could be used to identify patients who benefit from the addition of erlotinib. Using a combination of mutation, copy number, and gene expression analyses in the homogenous and clinically well-characterized CO-2016 cohort, I studied the effect of small patient groups on clinical outcome and erlotinib responsiveness. I sought to detect genetically and biologically distinct patient subgroups and identify molecular biological characteristics that can be used as biomarkers for pre-treatment screening. By subsequently modeling my findings *in vitro*, I was working toward reproducing the retrospective results from the CO-2016 cohort and gaining insight into their biological background.

2 Material & Methods

2.1 Material

2.1.1 Chemicals and Reagents

Reagent/Chemical	Supplier (Cat. No.)
2-Mercaptoethanol	AppliChem (A1108)
Albumin fraktion V	Carl Roth (8076)
Amersham ECL-Prime Western Blotting Detection Reagent	GE Healthcare (RPN2232)
Ammonium peroxydisulfate (APS)	Carl Roth (9592)
Ampicillin sodium salt	Carl Roth (HP62)
Ampure BeadsXP	Beckman Coulter (A63880)
cOmplete, Mini, EDTA-free Protease Inhibitor Cocktail	Roche (4693159001)
ddPCR Droplet Generator Oil for Probes	Bio-Rad (186-3005)
Dimethyl sulfoxide (DMSO)	Sigma Aldrich (41639)
Ethanol (99%)	Carl Roth (K928.1)
Ethylene diamine tetraacetic acid (EDTA)	Carl Roth (8043.1)
Fetal bovine serum (FBS)	Thermo Scientific (F2442)
GelRed	Biotium (41003)
GeneRuler 100 bp Plus DNA Ladder	Thermo Scientific (SM0322)
Glycine	Carl Roth (3790.1)
L-Buthionine-(S,R)-Sulfoximine	Cayman Chemicals (14484)
L-Glutamin (200mM)	Gibco (25030081)
N,N,N′,N′-Tetramethylethylenediamine (TEMED)	Sigma Aldrich (T9281)
Penicillin-Streptomycin (10,000 U/mL)	Gibco (15140122)
PhosStop	Roche (4906845001)
Ponceau S	Carl Roth (5938.1)
Protein Marker VI prestained	AppliChem (A8889)
RNase A	Qiagen (19101)
Tris-Pufferan	Carl Roth (4855.3)
Trypan Blue Solution	Thermo Scientific (15250061)
Trypsin-EDTA (0.5%)	Gibco (15400054)
UltraPure Agarose	Invitrogen (16500100)
UltraPure DNase/RNase-Free Distilled Water	Invitrogen (10977023)

2.1.2 Kits

Kit	Supplier (Cat. No.)
Agilent TapeStation D1000	Agilent (5067-5582/3)
Agilent TapeStation D1000 High Sensitivity	Agilent (5067-5584/5)
Agilent TapeStation Genomic DNA	Agilent (5067-5366/7)
Agilent RNA 6000 Nano Kit	Agilent (5067-1511)
Cell Proliferation Reagent WST-1	Roche (11644807001)
ddPCR Supermix for Probes (No dUTP)	Bio-Rad (1863023)
dNTP set 100mM solution	Thermo Fischer Scientific (R0181)
HiScribeTM T7 High Yield RNA Synthesis Kit	New England Biolabs (E2040S)
HotStar Taq Polymerase	Qiagen (203203)
KAPA Hyper Prep Kit	Roche (KK8504)
KAPA HiFi HotStart ReadyMix	Roche (KK2601)
Luna Universal qPCR Master Mix	New England Biolabs (M3003)
MiSeq Reagent Kit v2 (300-cycles)	Illumina (MS-102-2002)
MMLV reverse transcriptase	Invitrogen (28025021)
NEB Next Multiplex Oligos For Illumina (96)	New England BioLabs (E7600S)
NEB Next Ultra DNA Library Prep Kit for Illumina	New England BioLabs (E7370L)
NEBNext® Multiplex Oligos for Illumina (Index Primers Set 1)	New England Biolabs (E7335L)
NEBNext® Ultra™ II DNA Library Prep Kit for Illumina	New England Biolabs (E7645L)
Neon™ Transfektionssystem 10 µl-Kit	Thermo Fischer Scientific (MPK1096)
Pierce BCA protein Kit	Thermo Fischer Scientific (23227)
QIAamp DNA Mini Kit	Quiagen (51306)
QIAquick PCR Purification Kit	Quiagen (28106)
QuantiFluor ONE dsDNA dye	Promega (E4871)
QuantiFluor RNA System	Promega (E3310)
Random Primers	Invitrogen (48190011)
RNase-free DNase Set	Quiagen (79254)
RNeasy Mini Kit	Quiagen (74104)
SALSA MLPA EK5 reagent kit – 500 rxn – FAM	MRC Holland (EK5-FAM)
Sure Select XT Target Enrichment Kit	Agilent
Zymo Research DNA Clean & Concentrator-5 kit	Zymo Research (D4029)
RNA Clean & Concentrator-5	Zymo Research (R1013)

RNeasy Mini Kit	Qiagen (74104)
WT1 TaqMan SNP Genotyping assay	Thermo Fischer Scientific (4351379)

2.1.3 Oligonucleotides

2.1.3.1 PCR and sequencing primers

Name	Sequence	Name	Sequence
*ACVR1B*_chr12_52385733_F	TCCCAGGAGTCCATGAAGA	*ACVR1B*_chr12_52385733_R	TCAGGCTGATGGGAAAGC
*ACVR2A*_chr2_148657040_F	TCCAGGAACATTTTTGGTTTTT	*ACVR2A*_chr2_148657040_R	TATTGCCCTCACAGCAACAA
*APC*_chr5_112179738_F	TAGTGGGACTGTTGCTGCC	*APC*_chr5_112179738_R	TGGCGCTTAGGACTTTGG
*ARID1A*_chr1_27023140_F	CGGGGAAATGAAGGCAG	*ARID1A*_chr1_27023140_R	CGTTCCCGTTCGAGTTCT
*ARID2*_chr12_46233279_F	CCACAAGCTATAGAGCAAGTCC	*ARID2*_chr12_46233279_R	CCAAATTCCACTTTGGCCT
*ATM*_chr11_108117823_F	GGAAGTAGAGGAAAGTATTCTTCAGG	*ATM*_chr11_108117823_R	GGCTTTTTGTGAGAACACAGG
*BCORL1*_chrX_129168490_F	TGGAGCCAGCAGAAGACA	*BCORL1*_chrX_129168490_R	GGAGGGTCTCACCAGCATT
*BRAF*_chr7_140482957_F	AGCCAGGTAATGAGGCAGG	*BRAF*_chr7_140482957_R	CCTTTTTGGCCCATCCTT
*BRCA1*_chr17_41245251_F	GCTGCACACTGACTCACACA	*BRCA1*_chr17_41245251_R	TGCTGAAGACCCCAAAGA
*BRCA2*_chr13_32906781_F	CTGAAGTGGAACCAAATGATACTG	*BRCA2*_chr13_32906781_R	GGGTATTTTCTCCATCTGGG
*CDKN2A*_chr9_21994140_F	CCAAACAAAACAAGTGCCG	*CDKN2A*_chr9_21994140_R	TCGTGGTTCACATCCCG
*EGF*_chr4_110914413_F	CCCCACTTTTCTCTCCCGT	*EGF*_chr4_110914413_R	CCAACCTACCAGGGCAAA
*EGFR*_chr7_55273200_F	AGCACATTCGACAGCCCT	*EGFR*_chr7_55273200_R	CCGTGGTCATGCTCCAAT
*ERBB2*_chr17_37872060_F	ATAGTGGCGCCTACTCGCT	*ERBB2*_chr17_37872060_R	TTGGCAGTGTGGAGCAGA
*ERBB3*_chr12_56491000_F	GGGTCATCTCTGCAGCTTG	*ERBB3*_chr12_56491000_R	CACTTTTTCCTGGCTGGC
*ERBB4*_chr2_212578285_F	GGTTTGGTTGTGAGAGTGGG	*ERBB4*_chr2_212578285_R	TGCCTGCCCTAGTTCCAA
*FBXW7*_chr4_153247295_F	TATGCCCCTGCAACGTGT	*FBXW7*_chr4_153247295_R	GCCACTCTTAGGGTTTGGG
*GATA6*_chr18_19751322_F	CTCTTCCTCGTCCTCCTCCT	*GATA6*_chr18_19751322_R	GTCGAGGTCAGTGAACAGCA
*GLI3*_chr7_42005685_F	TTGCAGCTGCTGAGGCT	*GLI3*_chr7_42005685_R	AGTTCATGCCCCGAGGA
*GLI3*_chr7_42017192_F	AATGGCCAAGGGGAAAAC	*GLI3*_chr7_42017192_R	TGTGAGCACGAAGGTTGC
*GNAS*_chr20_57484421_F	TTTCGGTTGGCTTTGGTG	*GNAS*_chr20_57484421_R	AGAGGGACTGGGGTGAATG
*HUWE1*_chrX_53578298_F	TTCCTCCTGAATGGCTGG	*HUWE1*_chrX_53578298_R	TCTTTTCTGGCTGCCCTG
*KDM6A*_chrX_44929106_F	CAGGTGGACAACAAGGCA	*KDM6A*_chrX_44929106_R	CCATGGCTAGGACTGCAAA
*KMT2C*_chr7_151919692_F	TGGACAGTAAGGTGCAAGTGA	*KMT2C*_chr7_151919692_R	GGATGCATGCAGTTTGTCA
*KMT2C*_chr7_152027758_F	GTCATTTCACTGCTTTATTCCTGTC	*KMT2C*_chr7_152027758_R	TCTGCAGAAGAGGATGCTGA
*KMT2D*_chr12_49418395_F	GCATCACAAGCTCACCGTT	*KMT2D*_chr12_49418395_R	TTGGCCTAACCCCAAAAA
*MAP2K4*_chr17_12016647_F	TGACTTCGGCATCAGTGG	*MAP2K4*_chr17_12016647_R	GCAGTTTCATGGTAAAACCGT
*MAP2K7*_chr19_7976189_F	GATACGTCTTCTCCTCCCCC	*MAP2K7*_chr19_7976189_R	TGTTCTCCTCCCTCCGAGT
*MAPT*_chr17_44068882_F	TTTGAGTCAAGGTGGCGG	*MAPT*_chr17_44068882_R	CGGGGTTTTTGCTGGAAT

*MARK2*_chr11_63662808_F	GTCCAACATGATTCGGGG	*MARK2*_chr11_63662808_R	TGTCAACAGGGCTGGGA
*MSH2*_chr2_47630410_F	CGCGCATTTTCTTCAACC	*MSH2*_chr2_47630410_R	CGTGCGCCGTATAGAAGTC
*MYC*_chr8_128752653_F	TGTGGCTCTTTGGGGAGA	*MYC*_chr8_128752653_R	GAGCCTGCCTCTTTTCCA
*MYCBP2*_chr13_77671652_F	TGGATGTGACTCCCCACA	*MYCBP2*_chr13_77671652_R	CCTGGTTCCTGTGCAGTTC
*NF2*_chr22_30057302_F	GAGAACAGACTGACCCCCAA	*NF2*_chr22_30057302_R	CCCTCAAAGCCTGGGAAT
*NRG1*_chr8_32620810_F	ACAAGGCACACAGATCCAAA	*NRG1*_chr8_32620810_R	GGATTCGGAAGGGAAAGAAG
*PALB2*_chr16_23637699_F	TTTCCCACTGCCAAGCAT	*PALB2*_chr16_23637699_R	CTTTGCATAAAACAGCACTCG
*PBRM1*_chr3_52649473_F	GGCCCAGCTCTTTCCTTT	*PBRM1*_chr3_52649473_R	TGAGGCTGGTTTTGGTTTTC
*PBRM1*_chr3_52712614_F	AAATCCCCCGCAACTGTA	*PBRM1*_chr3_52712614_R	TGTCTACACTGGCTTTCTGTTCA
*PIK3CA*_chr3_178916929_F	TGACCTTCGGCTTTTTCAAC	*PIK3CA*_chr3_178916929_R	CGAAGGTATTGGTTTAGACAGAAA
*PIK3CG*_chr7_106508418_F	AAGAAGGGGCAGTGGTACG	*PIK3CG*_chr7_106508418_R	TCATAGCCAATCAGCGCC
*PLCG2*_chr16_81888157_F	GAGCAAAAGCAGTTCGCC	*PLCG2*_chr16_81888157_R	ACATCAGTGAGCAGGCAGG
*PLXNB2*_chr22_50718981_F	CAGCATCAGCTTGGGGTT	*PLXNB2*_chr22_50718981_R	TCATCCACACCCTGGAGAA
*PREX2*_chr8_68992692_F	TTGAGGTTATGGGTGCAGG	*PREX2*_chr8_68992692_R	CCTCCATTTCCTCATCCG
*PRKCG*_chr19_54403957_F	TGGATGCTGAGGGACACA	*PRKCG*_chr19_54403957_R	CACCCGTGGTGAATACTGG
*PRSS1*_chr7_142458429_F	TCTCCCTTCCCATCTCCA	*PRSS1*_chr7_142458429_R	GGCCCCACACTTACGACTT
*RBM10*_chrX_47039684_F	GGAAATCGGGCAATGGA	*RBM10*_chrX_47039684_R	TGGGAAGGCACACAGGT
*RBM6*_chr3_50098589_F	TCCTCAGGTGACTATAAGGGTGA	*RBM6*_chr3_50098589_R	CACTGTTGTCCTTGTCGATTTG
*RNF43*_chr17_56435189_F	TGGGGAAAAATCTGGCAA	*RNF43*_chr17_56435189_R	AGTTGACGCCTCCAGCAT
*RNF43*_chr17_56435408_F	GCTGCTGAGTTGGATCTGG	*RNF43*_chr17_56435408_R	AAACCCAGGTTTCCAGCC
*ROBO1*_chr3_78734899_F	ACTGGCACCCTCTCTGGAT	*ROBO1*_chr3_78734899_R	CAGACTGGCGACCTCACAA
*ROBO2*_chr3_77684143_F	TCGACCTACCAGCCCATT	*ROBO2*_chr3_77684143_R	TGGAGCACGTCTCTTCAGC
*SETD2*_chr3_47164003_F	TGCAGCATGCAGGTATCA	*SETD2*_chr3_47164003_R	AAATGGATCTCCTGGGGC
*SF3B1*_chr2_198267371_F	TGTGGCAAGATGGCACAG	*SF3B1*_chr2_198267371_R	CATGGATGAGTATGTCCGTAACA
*SLIT2*_chr4_20525647_F	CTCTTCGGGCCATTCAAA	*SLIT2*_chr4_20525647_R	TCCGCTAGCCACTTGAGA
*SMAD3*_chr15_67477192_F	TGCGGCTCTACTACATCGG	*SMAD3*_chr15_67477192_R	TAGGCATGCACTCGGAGA
*SMAD4*_chr18_48591919_F	GCTCCTGAGTATTGGTGTTCC	*SMAD4*_chr18_48591919_R	TGGCTTCTGTCCTGTGGA
*SMARCA2*_chr9_2104109_F	GCCTCAGGGAAGTTTGAGC	*SMARCA2*_chr9_2104109_R	TGGCTTCCGAGCCTAATG
*SMARCA4*_chr19_11123707_F	CAACCTGAACGGCATCCT	*SMARCA4*_chr19_11123707_R	AACACAGGAGCGTGTGAGC
*TGFBR1*_chr9_101900164_F	TCAGTTTTCTGGGTCACTCATT	*TGFBR1*_chr9_101900164_R	CGCCACTTTCCTCTCCAA
*TGFBR2*_chr3_30713626_F	GCAGTTTGAGACAGTGGCAG	*TGFBR2*_chr3_30713626_R	TAGGTTGCCCTTGGCGT
*TGFBR2*_chr3_30729962_F	GGCTGCACATGCCATTCT	*TGFBR2*_chr3_30729962_R	TACCTGGTGGTTGAGCCAG
*TLE4*_chr9_82188691_F	CCGAATCCTGTGATCGGA	*TLE4*_chr9_82188691_R	CTTTCTGCCCACCGAGAA
*TP53*_chr17_7578469_F	AACCAGCCCTGTCGTCTCT	*TP53*_chr17_7578469_R	TTGCCAACTGGCCAAGAC
*TP53BP2*_chr1_223987624_F	CCTGTTTCTTCTGCTCTCCC	*TP53BP2*_chr1_223987624_R	AGTCCAATGCCCCACCTT
*U2AF1*_chr21_44524456_F	AAAGGCAAACAAACCTGGC	*U2AF1*_chr21_44524456_R	GCTGCTGACATATTCCATGTG

PDAC_*ACVR2A*_148683680_F	AGGAGGAAATTGGCCAGC	PDAC_*ACVR2A*_148683680_R	TGCAGAAGAAAGAGAAATGTGC
PDAC_*APC*_112175394_F	ATCAGCCAGGCACAAAGC	PDAC_*APC*_112175394_R	AGCTGGCAATCGAACGAC
PDAC_*ARID1A*_27089711_F	GCATCCCACCTTATGGCA	PDAC_*ARID1A*_27089711_R	CAACATGCATGGCGACAG
PDAC_*ARID1A*_27106618_F	GCTCCGGGAAAACACCTT	PDAC_*ARID1A*_27106618_R	GAAAGGACGGCATTGGG
PDAC_*BCORL1*_129147548_F	ATGCATTCCAGGTTCCCC	PDAC_*BCORL1*_129147548_R	CCAAGACTGAGGCCGAAA
PDAC_*BRCA1*_41243728_F	CAAGTCTTCCAATTCACTGCAC	PDAC_*BRCA1*_41243728_R	CACCGTTGCTACCGAGTGT
PDAC_*BRCA2*_32911709_F	ATCAAAAGAAACTGAGCAAGCC	PDAC_*BRCA2*_32911709_R	AATTTCTGCCTTTTGGCTAGG
PDAC_*BRCA2*_32929256_F	GCAGTTTCAGGACATCCATTT	PDAC_*BRCA2*_32929256_R	GAGCCATGTCCATCAATGTTT
PDAC_*BRCA2*_32944580_F	GGCAGTTCTAGAAGAATGAAAACTC	PDAC_*BRCA2*_32944580_R	GCGATGATAAGGGCAGAGG
PDAC_*BRCA2*_32968951_F	TGCCCCTTTCGTCTATTTG	PDAC_*BRCA2*_32968951_R	GGCCCTCTTTTGGACTAGC
PDAC_*BRCA2*_32969065_F	AAGAGGGCCACTTTCAAGA	PDAC_*BRCA2*_32969065_R	CCTGAGGTTCATGGGCA
PDAC_*ERBB3*_56481669_F	AGCCCTAACAGCCATGCTT	PDAC_*ERBB3*_56481669_R	GGCCACATCCCAATTTCA
PDAC_*FBXW7*_153245440_F	TGTCCTGTTTTGATATCCCAGA	PDAC_*FBXW7*_153245440_R	CCATGTGGTGAGTGGATCTCT
PDAC_*GLI3*_42005312_F	TCCACTGAATGGGCAGGT	PDAC_*GLI3*_42005312_R	TTCCCAGAACCAAGCAGG
PDAC_*GLI3*_42007392_F	CATCCATTTGGTCCCTGC	PDAC_*GLI3*_42007392_R	ACAGTCCCCCATCAGCAA
PDAC_*GNAS*_57415795_F	CAAGCCCGAGGACAAAGA	PDAC_*GNAS*_57415795_R	GACGGCGTCCTCCATTAGT
PDAC_*GNAS*_57484420_F	TTTCGGTTGGCTTTGGTG	PDAC_*GNAS*_57484420_R	AGAGGGACTGGGGTGAATG
PDAC_*KDM6A*_44922710_F	TGACAGATGAGACCAACAGGA	PDAC_*KDM6A*_44922710_R	CAGAAAAGGGTCCATTGGC
PDAC_*KDM6A*_44941820_F	GGTGGAAAGGATTGCCAG	PDAC_*KDM6A*_44941820_R	GAGACACAGTATTAGAAACATGCC
PDAC_*KMT2C*_151860763_F	TGGGCATTCTAACAGGGC	PDAC_*KMT2C*_151860763_R	AATGGCCCCACCTACCAT
PDAC_*KMT2C*_151878005_F	CCTTGGGAGTGCATTGGA	PDAC_*KMT2C*_151878005_R	TTTAGCCGTGTTTCCCCA
PDAC_*KMT2C*_151945115_F	ACCCTACCTGTTTGGACCG	PDAC_*KMT2C*_151945115_R	TGTGTCTTCCTCCCCAACA
PDAC_*KMT2D*_49436599_F	AAAGAGCAACTGGCCCATC	PDAC_*KMT2D*_49436599_R	TGGAATCCAAGGACCTGC
PDAC_*KMT2D*_49437175_F	AGGGACAGAGGCAAGCAA	PDAC_*KMT2D*_49437175_R	GGAGACCTGTTCTGGGCAT
PDAC_*KMT2D*_49439899_F	ACAAACGGAGGTGGCTGA	PDAC_*KMT2D*_49439899_R	ATGCATGCAGGCTGTGAG
PDAC_*KMT2D*_49446487_F	TTGGCATGCAACGTACAGA	PDAC_*KMT2D*_49446487_R	TAAGGTGGGCAGTGCCAT
PDAC_*KRAS*_25380275_F	AAAGAAAGCCCTCCCCAG	PDAC_*KRAS*_25380275_R	AAAGGTGCACTGTAATAATCCAGAC
PDAC_*MAP2K7*_7975197_F	CAGGCAGAAATCAACGACC	PDAC_*MAP2K7*_7975197_R	ATTTGCTGCACAGGCAGG
PDAC_*MYCBP2*_77672687_F	AAGGTTCTGTGAGGGGCA	PDAC_*MYCBP2*_77672687_R	CCCATCGGGTGCTAGTTCT
PDAC_*PBRM1*_52613086_F	TGCCAGGGGAAGGACTTT	PDAC_*PBRM1*_52613086_R	ATGTGGCTGAAGGTTGGC
PDAC_*PBRM1*_52623100_F	GGAGCAAGGAGAAGTCAAACA	PDAC_*PBRM1*_52623100_R	TCGCACATACAGCCAGGA
PDAC_*PBRM1*_52643657_F	TCCTCATCTCCCTCCAGGT	PDAC_*PBRM1*_52643657_R	AGCCCATGGACATGGAAA
PDAC_*PBRM1*_52643767_F	CAAAGACTCCGGCTCATTGT	PDAC_*PBRM1*_52643767_R	CCGCCTCAGTGCCATATT
PDAC_*PIK3CA*_178927451_F	GGTGGAATGAATGGCTGAA	PDAC_*PIK3CA*_178927451_R	TCAATCAGCGGTATAATCAGGA
PDAC_*PIK3CG*_106508260_F	GCAACGTGGAGCAGATGA	PDAC_*PIK3CG*_106508260_R	TGCGTGGCCTTCCAGTA
PDAC_*PLCG2*_81891943_F	GGCAGGGCTTTGTAAGCA	PDAC_*PLCG2*_81891943_R	CGCAAATCAACAGGCAAA

PDAC_*PLXNB2*_50716674_F	TCTTGGACTTGCCCTCGT	PDAC_*PLXNB2*_50716674_R	TGTCTGCCCTGGACACAT
PDAC_*PLXNB2*_50718100_F	GATGCCCAAGCTCCTTAGC	PDAC_*PLXNB2*_50718100_R	GCCATCAAACATCAGGTGG
PDAC_*PLXNB2*_50728692_F	ATTGCTGGCCACGAAAGA	PDAC_*PLXNB2*_50728692_R	TGCCATGAGGCTGAGATG
PDAC_*RBM6*_50005237_F	CACAGTTGGACTTCAGGGG	PDAC_*RBM6*_50005237_R	TGGAGGAGCATCCCTACCT
PDAC_*RBM6*_50006146_F	TCTGGACTTTCTTGGGCG	PDAC_*RBM6*_50006146_R	AGCAATCCACCCCATCAA
PDAC_*ROBO1*_78656164_F	TTGGCACCACTACAGGTCG	PDAC_*ROBO1*_78656164_R	GGTGAATGTGGCTCCAATG
PDAC_*ROBO1*_78737884_F	TGGCAACCATCTTTGCTG	PDAC_*ROBO1*_78737884_R	CCTCCTTCAGAACCTCCACA
PDAC_*ROBO1*_78987969_F	AGATCCACTCGGCAGCAA	PDAC_*ROBO1*_78987969_R	TCCACCTCGCATTGTTGA
PDAC_*SETD2*_47164859_F	TGTATGGCCGAGAATAGCG	PDAC_*SETD2*_47164859_R	TCGGTATGTATCTTCCCGATG
PDAC_*TGFBR2*_30732969_F	AACAAGGTCAGCAGGCCA	PDAC_*TGFBR2*_30732969_R	TCCAGATGCTCCAGCTCA
PDAC_*U2AF1*_44524456_F	AAAGGCAAACAAACCTGGC	PDAC_*U2AF1*_44524456_R	GCTGCTGACATATTCCATGTG
PDAC_*CDKN2A*_21970952_F	ACCTTCCGCGGCATCTAT	PDAC_*CDKN2A*_21970952_R	TTCCTGGACACGCTGGT
PDAC_*CDKN2A*_21974695_F	TGCAAACTTCGTCCTCCA	PDAC_*CDKN2A*_21974695_R	AGCAGCATGGAGCCTTC
PDAC_*CDKN2A*_21974799_F	AAAGAGGAGGGGCTGGCT	PDAC_*CDKN2A*_21974799_R	CACCTCCTCTACCCGACCC
PDAC_*TP53*_7576852_F	TTTGAGGCATCACTGCCC	PDAC_*TP53*_7576852_R	CACTGCCCAACAACACCA
PDAC_*TP53*_7577120_F	CCCCTTTCTTGCGGAGAT	PDAC_*TP53*_7577120_R	GACAAGGGTGGTTGGGAGT
PDAC_*TP53*_7577538_F	TGTGCAGGGTGGCAAGT	PDAC_*TP53*_7577538_R	CGCACTGGCCTCATCTT
PDAC_*TP53*_7578205_F	AACCACCCTTAACCCCTCC	PDAC_*TP53*_7578205_R	TTGCTCTTAGGTCTGGCCC
PDAC_*TP53*_7578538_F	TCAACCCACAGCTGCACA	PDAC_*TP53*_7578538_R	TAGTGGGTTGCAGGAGGTG
PDAC_*TP53*_7579539_F	AGAAGATGACAGGGGCCA	PDAC_*TP53*_7579539_R	ATCTACAGTCCCCCTTGCC
PDAC_*ATM*_108098366_F	ACAGACAGTGATGTGTGTTCTGA	PDAC_*ATM*_108098366_R	ACTACCTTTCGTTCTGTAGCTCT
PDAC_*ATM*_108098600_F	AAGCGCCTGATTCGAGATCC	PDAC_*ATM*_108098600_R	AGGAAGCAAAGATAAATGTTAAGACT
PDAC_*ATM*_108117799_F	TTTTGGATTACAGGTGCTTATGAATCA	PDAC_*ATM*_108117799_R	TGTGACATGACCTACTTACTGTACC
PDAC_*ATM*_108164163_F	AGGATAATGAAAACCTCTATATCACGA	PDAC_*ATM*_108164163_R	GCTAGAGCATTACAGATTTTTGAAAA
PDAC_*ATM*_108178700_F	TTTTCCGATGCTGTTTGGA	PDAC_*ATM*_108178700_R	CACTTTGCAACACCTTCACC
PDAC_*ATM*_108183190_F	AAAGAGGTGTTCTTGTGACAAACA	PDAC_*ATM*_108183190_R	TCACATGGCATCTGTACAGTGT
PDAC_*ATM*_108186796_F	ACACGAAGCAATGTGGGG	PDAC_*ATM*_108186796_R	TGGCTGTGTAAATATCCACCAA
PDAC_*ATM*_108199794_F	TGGTTGTGTTTTCTTGAAGGC	PDAC_*ATM*_108199794_R	CCTCTTTGGCTCTTTTCAGG
PDAC_*ATM*_108202221_F	TTGTGCTAATAGAGGAGCACTGTC	PDAC_*ATM*_108202221_R	AATCCTAGGCCTCCCATCA
PDAC_*ATM*_108203492_F	ACTGGCCTTAGCAAATGCAA	PDAC_*ATM*_108203492_R	AGATGGCATCTTAATAACACAAGATT
PDAC_*ATM*_108216476_F	TCCTGTTCATCTTTATTGCCC	PDAC_*ATM*_108216476_R	TTTGGCAAACATCCATGAAG
PDAC_*SMAD4*_48584709_F	GGATTTACTGGTCAGCCAGC	PDAC_*SMAD4*_48584709_R	TTTTGGTGGTGAGGCAAA
PDAC_*SMAD4*_48593532_F	TGACCACGCGGTCTTTGT	PDAC_*SMAD4*_48593532_R	TTCCATTCCTTCCACCCA
PDAC_*SMAD4*_ 48581161_F	TAAGGTTAGTAATTTACACTGTAA	PDAC_*SMAD4*_ 48581161_R	TGAATGTCCTTCAGTGGA
PDAC_*SMAD4*_48573530_F	TGTCTATTACGAATACACCAACAAGT	PDAC_*SMAD4*_48573530_R	ACACATTTACTAGGATGAGCTCC
PDAC_*SMAD4*_48575085_F	ACTGAGTTGGTAGGATTGTGAGG	PDAC_*SMAD4*_48575085_R	TTTAAGTCAAACGCATACTGACAA

PDAC_*SMAD4*_48575161_F	TCCTCATGTGATCTATGCCCG	PDAC_*SMAD4*_48575161_R	TCCTTTTCCCTTTATGTTTCTTAGGA
PDAC_*SMAD4*_48575210_F	TGGAGGTGGCCTGATCTT	PDAC_*SMAD4*_48575210_R	GAAACACTATTGAGATCCTTTTCCC
PDAC_*SMAD4*_48581236_F	TGAAGGACATTCAATTCAAACCATCC	PDAC_*SMAD4*_48581236_R	TAAGCAATGATTAGGAAAACTCAACT
PDAC_*SMAD4*_48584561_F	TTCTTGTTCCTCTAGGTCAGCC	PDAC_*SMAD4*_48584561_R	TGCTTTTATAAAGGCTGCCTACT
PDAC_*SMAD4*_48591821_F	CTGTGTTGTGGAGTGCAAGTG	PDAC_*SMAD4*_48591821_R	AGCGATCTCCTCCAGAAGGG
PDAC_*SMAD4*_48591914_F	CCTGAGTATTGGTGTTCCATTGC	PDAC_*SMAD4*_48591914_R	TCTGACTATACAATCAATACCTTGCT
PDAC_*SMAD4*_48591977_F	TGCCCTATTGTTACTGTTGATGGA	PDAC_*SMAD4*_48591977_R	ACACATATAATGTACATGGGAAAACA
PDAC_*SMAD4*_48593404_F	AAGCATGCTATACAATCTGAACTAAA	PDAC_*SMAD4*_48593404_R	ACGCCCAGCTTCTCTGTCTA
PDAC_*SMAD4*_48593476_F	GCAAAGGTGTGCAGTTGGAA	PDAC_*SMAD4*_48593476_R	GAACAGCATCTCCAGGTGCA
PDAC_*SMAD4*_48603039_F	TGAGTTTTAAATAAGTCAGGCATTGGT	PDAC_*SMAD4*_48603039_R	TCCACCTACTGATCCTGGGC
PDAC_*SMAD4*_48603130_F	CATCGACAGATGCAGCAGC	PDAC_*SMAD4*_48603130_R	TGCTCAAAGAAACTAATCAACTGAGT
PDAC_*SMAD4*_48604684_F	TCACCCTGTCCCTCTGATGT	PDAC_*SMAD4*_48604684_R	TTCGTCTAGGAGCTGGAGGG
PDAC_*SMAD4*_48604767_F	GGGACCGGATTACCCAAGAC	PDAC_*SMAD4*_48604767_R	GTTAAGGGCCCCAACGGTAA
PDAC_*CDKN2A*_21971176_F	ACCAGCGTGTCCAGGAA	PDAC_*CDKN2A*_21971176_R	GCTTCCTTTCCGTCATGC
PDAC_*TP53*_7577040_F	AACTGCACCCTTGGTCTCC	PDAC_*TP53*_7577040_R	GGGACGGAACAGCTTTGA
PDAC_*TP53*_7578406_F	AACCAGCCCTGTCGTCTCT	PDAC_*TP53*_7578406_R	TTGCCAACTGGCCAAGAC
PDAC_*ATM*_108124695_F	TGTGGTATAGAAAAGCACCAGTC	PDAC_*ATM*_108124695_R	TGCCCCTATTTCTCCTTCC
*SMAD4*_ex3_M13_F	CCACCATGCCCAGTCACTTT	*SMAD4*_ex3_M13_R	AAAGTCGCGGGCTATCTTCC
*SMAD4*_Ex6+7_F	TGCCCATCTTTATAGTTGTGCA	*SMAD4*_Ex6+7_R	CAAAATCACAGGATGAATAAAAGATGT
*SMAD4*_Ex9_F	TCCAGTTGTTTTGGGTGCAT	*SMAD4*_Ex9_R	ACAAAAACACCGACAATTAAGATGGA

2.1.3.2 qPCR primers

Name	**Sequence**
*BAD*_RT-PCR_F	TTCCAGATCCCAGAGTTTGAGC
*BAD*_RT-PCR_R	CCCATCCCTTCGTCGTCCT
*CBP*_RT-PCR_F	GAATCAGCTCTTCCGACTTC
*CBP*_RT-PCR_R	TGCCAGCCTTTCCTTACA
*KIT*_RT-PCR_F	AGTGCATTCAAGCACAATGG
*KIT*_RT-PCR_R	CATGCCAGCTACGATTACGA
*MAPK9*_RT-PCR_F	ATGACCCCTTACGTGGTGACA
*MAPK9*_RT-PCR_R	CATGATGCAACCCACTGACC
*MECOM*_RT-PCR_F	tgaggatgactatgaagaaaccagt
*MECOM*_RT-PCR_R	gcagaaagtccacttttatattcttcc
*PLD1*_RT-PCR_F	GCCTATGGAAGGTGGGACGACA
*PLD1*_RT-PCR_R	CATTGCGGCAGGTGGGAGG

*RAC1*_RT-PCR_F	CCTGATGCAGGCCATCAAG
*RAC1*_RT-PCR_R	TGAATCTGGGCTTATGGGATACA
*SMAD4*_RT-PCR_F	CGCTTTTGTTTGGGTCAACT
*SMAD4*_RT-PCR_R	CCCAAACATCACCTTCACCT

2.1.3.3 Oligonucleotides for Cas9-sgRNA RNPs

Name	**Target exon**	**Ordered Oligonucleotide**
*SMAD4*_ex3_T1	3	AGACGGGCATAGATCACATGAGG
*SMAD4*_ex6_T1	6	GTCAGCCTGCCAGTATACTGGGG
*SMAD4*_ex9_T1	9	GATGGATACGTGGACCCTTCTGG

2.1.3.4 siRNA and ddPCR Kits

Name	**Supplier (Cat. No.)**
ddPCR™ *KRAS* G12/G13 Screening Kit	Bio-Rad (863506)
SiRNA *MAPK9* (Assay ID: s11159)	Thermo Scientific (4390824)

2.1.4 Western blot antibodies

Target	**Dillution**	**Supplier (Cat. No.)**
JNK2 (MAPK9)	1:1000	Cell Signalling (4672)
EGFR	1:1000	Cell Signalling (2232)
c-Jun (60A8)	1:1000	Cell Signalling (9165)
Phospho-c-Jun (Ser63) (54B3)	1:1000	Cell Signalling (2361)
SMAD4 (B-8)	1:1000	Santa Cruz (sc-7966)
GAPDH	1:500	Genetex (GT239)
Vincullin	1:1000	Cell Signalling (4650)

2.1.5 Buffers and Media

2.1.5.1 Purchased buffer and media

Solution	Supplier (Cat.No)
PBS	Gibco (10010023)
Rotiphorese® 50x TAE Puffer	Carl Roth (CL86.1)
RPMI	Gibco (21875034)
DMEM	Gibco (11965092)
DMEM/F12	Gibco (11320033)

2.1.5.2 Prepared buffer and media

Buffer	Components
10x TBS	1.54 M NaCl + 1.3 M Tris (pH 7.5)
10x Transfer buffer	2 M Glycin + 0.25 M Tris
10x Electrophoresis buffer	2 M Glycin + 0.25 M Tris + 35 mM SDS
Stacking Gel buffer	0.5 M Tris (pH 6.89)
Resolving Gel buffer	2 M Tris (pH 8.0)
Lysis buffer NP-40	50 mM Tris + 150 mM NaCl + 1 mM EDTA + 1 % NP-40
PANC1 Culture Medium	DMEM + 10 % FBS + 4 mM Glutamine + 1x Penicillin / Streptomycin
BxPC3 Culture Medium	RPMI + 10 % FBS + 2 mM Glutamine + 1x Penicillin / Streptomycin
Capan1/2 Culture Medium	RPMI + 20 % FBS + 2 mM Glutamine + 1x Penicillin / Streptomycin
HPAC Culture Medium	DMEM/F12 + 10 % FBS + 2 mM Glutamine + 1x Penicillin / Streptomycin

2.1.6 Devices and Supplies

Name	Supplier
2100 Bioanalyzer Instrument	Agilent
4200 TapeStation System	Agilent
AlphaImager	Protein Simple
Applied Biosystems StepOne Plus	Thermo Fisher Scientific
Covaris E220 Focused-ultrasonicator	Covaris
Infinite 200 PRO microplate readers	Tecan
MiSeq Sequencer	Illumina
NanoDrop 1000	Thermo Fisher Scientific
Neon Transfection System	Thermo Fisher Scientific
Nitrocellulose membrane Amersham Protran Premium	GE Healthcare
Quantus Fluorometer	Promega
TapeStation 4200 Loading Tips	Agilent

2.1.7 Software

Software	Version
R	3.4.0 – 3.6.2
R Studio	1.2.5033
Cytoscape	3.4.0
NSolver	4.0
Finch TV	1.4
StepOne Software	2.3
ImageJ	1.8.0
Coffalyser.Net	140721.1958
SPSS	25
IGV	2.7.2

2.1.8 Cell Lines and Primary Material

2.1.8.1 Cell lines

Cell line	Tissue of origin	Supplier (Cat.No)
PANC-1	Primary PDAC	DSMZ (ACC783)
BxPC-3	Primary PDAC	DSMZ (ACC760)
Capan-1	Liver metastasis of PDAC	DSMZ (ACC244)
Capan-2	Primary PDAC	DSMZ (ACC245)
HPAC	Primary PDAC	ATCC (CRL-2119)

2.1.8.2 Primary material

All patients were part of the CONKO-005 study. Formalin-fixed paraffin embedded (FFPE) resected pancreatic tissue was collected from 331 adult patients (age, 24-82 years) by the CONKO-005 study groups. The 293 patients subsequently analysed in this thesis are referred to as CO-2016 cohort.

Additionally, DNA from 20 tonsillectomy FFPE tissue samples was extracted as non-paired normal tissue.

2.2 Methods

2.2.1 Clinical study design and patient sample collection

All patients were enrolled in the CONKO-005 study, an open-label, multi-center, randomized phase III trial investigating the addition of erlotinib to gemcitabine compared to gemcitabine only as adjuvant therapy. Following R0 resection, 436 PDAC patients were enrolled and randomly assigned to either the experimental arm (Gem+Erlo [n=219]) or the standard treatment arm (Gem [n=217]). Patients in the Gem+Erlo arm were treated with gemcitabine 1,000 mg/m^2 on days 1, 8, and 15 every four weeks in combination with erlotinib, of which 100 mg was administered orally once per day on days 1 through 28 every 4 weeks. Those in the gemcitabine arm alone were treated with the same dose and interval of gemcitabine monotherapy, both for 24 weeks of therapy[142]. Tumour material was collected from 331 adult patients (aged 24–82 years) by the CONKO-005 study groups. Written consent was obtained in accordance with the Declaration of Helsinki and with ethical approval obtained from the local ethics committee.

In the central biobank of the Charité (ZeBanC), the paraffin-embedded blocks from preintervention surgical resection were assessed visually for determination of highest tumor content areas. Samples with a tumor content below 10% were excluded from this study. Samples with a tumor content below 50% were macrodissected to enrich the tumor cell fraction; those with a tumor content above 50% were directly processed. DNA and RNA were extracted by the ZeBanC using the half-automated Maxwell system (Promega) with the Maxwell 16 FFPE Tissue LEV DNA Purification Kit (Promega) and the RSC RNA FFPE Kit (Promega) according to the manufacturer's instructions.

2.2.2 Cell Biology

2.2.2.1 Cell culture

2.2.2.1.1 Cultivation and passaging

A panel of five PDAC cell lines was chosen to reflect the heterogeneity of PDAC. Cell lines used were purchased for DSZM and ATCC and are listed in section 2.1.8.1. All cells were cultured at 37°C with 5% CO_2 and 100% humidity in cell-specific media, as indicated in section 2.1.5.1. For passaging, medium was removed. Cells were rinsed once with Ca^{2+}-free PBS and incubated with an appropriate amount of Trypsin/EDTA at 37°C until detachment could be observed. Afterwards, cells were rinsed with medium and transferred to a new cell culture flask.

2.2.2.1.2 Storage of cells

For long-term storage, cells were detached and counted using a Neubauer cell-counting chamber. Next, 1x10^6–5x10^6 cells were centrifuged at 400xg for 5 minutes and resuspended

in 1 ml freezing medium containing 90% FCS and 10% dimethylsulfoxid (DMSO). Cells were then transferred to cryogenic vials and placed in a freezing container. The cell suspension was set freeze overnight at -80°C and was then transferred to liquid nitrogen. For subsequent thawing, cells were warmed in 37°C water bath, resuspended in 12 ml cell culture medium, and transferred to a 75 cm^2 cell culture flask. Following overnight adhesion, the medium was exchanged with fresh medium to remove remaining DMSO.

2.2.2.2 Generation of CRISPR/Cas9 *SMAD4* knock-out single cell clones using the Cas9-sgRNA ribonucleoprotein (RNPs)-based delivery system

To generate stable *SMAD4* single-cell knockout clones, I chose the Cas9-sgRNA ribonucleoprotein (RNPs)-based delivery system. This method, while based on the traditional CRISPR/Cas9 gene editing method, increases disruption efficiency, simplifies transfection set-up, and lowers turnaround times from those of plasmid-based CRISPR/Cas9. In so doing, rather than perform a transfection with a plasmid containing the CRISPR/Cas9 sequences, the final components of CRISPR (Cas9 protein and *in vitro* transcribed sgRNA) were pre-complexed and delivered to the target cells via electroporation[143] (Figure 2.1).

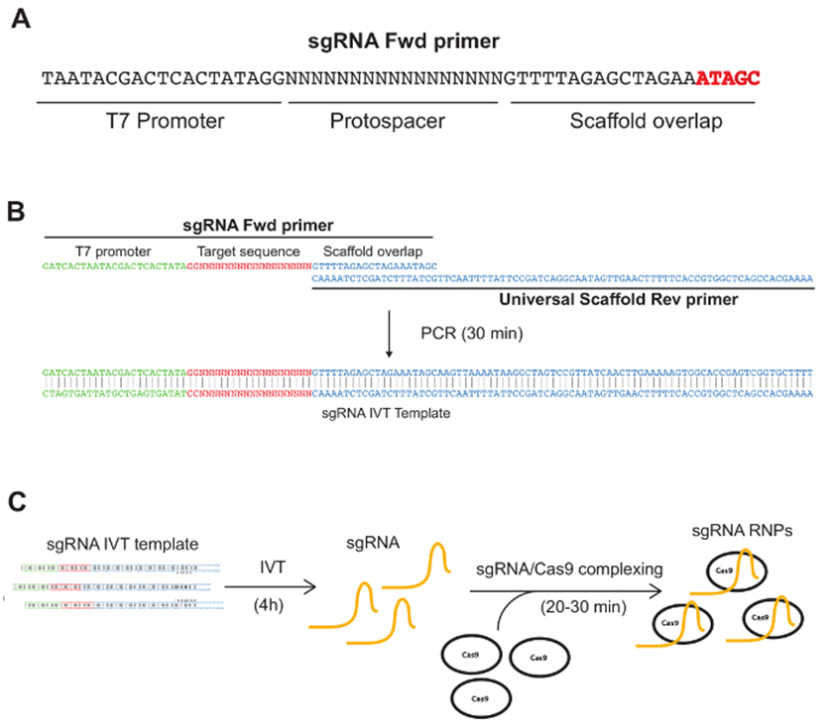


Figure 2.1: Mechanism of action for the Cas9-sgRNA ribonucleoprotein (RNPs)-based delivery system. A) Depiction of the sgRNA forward primer, an oligonucleotide used for in vitro generation of the guide RNAs. The sgRNA must contain the T7 promoter, the protospacer sequence, and an overlap sequence with the sgRNA scaffold as well as a 'ATAGC' sequence (red). **B)** Shema of the *in vitro* transcription via overlap PCR performed using a sgRNA forward primer and a universal scaffold reverse primer. The overlap between both primers allows for extension and amplification via PCR. **C)** The final sgRNA template was amplified and transcribed via in vitro transcription (IVT) and pre-complexed with Cas9 protein. The complex can then be transferred into target cells using electroporation. Graph from Brunetti et al.[144]**Cas9-sgRNA RNP generation**

The first step in the generation of Cas9-sgRNA RNP complexes was the *in vitro* transcription of the guide RNAs. To this end, oligonucleotides were designed with a T7 promotor sequence, the target sequence, and an overlap sequence for the scaffold primer (Figure 2.1 B). I designed three oligonucleotides targeting regions in exon 3, exon 6, and exon 9 of *SMAD4* (0). To obtain the sgRNA DNA template for in vitro transcription, an overlapping PCR was performed as described by Brunetti et al[144] using a sample mix containing a high-fidelity DNA polymerase (Table 2.1) and running the below PCR program (Table 2.2).

Table 2.1: Composition of overlapping PCR master mix.

Component	Volume
2x KAPA HiFi HotStart ReadyMix	10 µl
sgRNA oligonucleotide (10 µM)	2 µl
Universal reverse scaffold primer (10 µM)	2 µl
Nuclease-free H_2O	6 µl

Table 2.2: Cycler parameters used for overlapping PCR.

Step	Temperature	Time	Cycles
Initial denaturation	95 °C	3 min	
Denaturation	98 °C	5 s	6x
Annealing	60 °C	5 s	
Extension	72 °C	10 s	
Final extension	72 °C	1 min	
Hold	4 °C	∞	

The PCR products were subsequently purified with the Zymo Research DNA Clean & Concentrator-5 kit according to the manufacturer's instructions and eluted with an 11.5 µl elution buffer. DNA concentration was measured with the NanoDrop1000 spectrophotometer and 200–480 ng of the purified PCR product was used for *in vitro* transcription with the HiScribe™ T7 High Yield RNA Synthesis Kit. The components of the kit were mixed as previously described (Table 2.3) and the reaction was incubated at 37°C for 16 hours.

Table 2.3: Reaction mix for in vitro transcription using the HiScribe™ T7 High Yield RNA Synthesis Kit.

Component	Volume

10x reaction buffer	1 µl
ATP (100 mM)	1 µl
CTP (100 mM)	1 µl
GTP (100 mM)	1 µl
TTP (100 mM)	1 µl
purified PCR product	4 µl
T7 RNA polymerase enzyme mix	1 µl

In a final step, each RNA was filled to 50 µl and purified using the Zymo Research RNA Clean and Concentrator-5 kit following the manufacturer's instructions. The RNA was eluted in 35 µl of nuclease free water and quantified using a NanoDrop1000 spectrophotometer.

2.2.2.2.2 Electroporation

To introduce the Cas9-sgRNA RNP complex into PANC1 cells, I used electroporation with the Neon transfection system. In a first step, the synthetic Cas9 protein and the *in vitro* transcribed sgRNAs were thawed on ice, mixed in a molar ratio of 1:10 (3 pmol recombinant Cas9 nuclease, 0.5 µg, were mixed with 30 pmol guide RNA per replicate) and incubated at room temperature for 30 minutes. In the meantime, a 24-well plate was filled with 500 µl antibiotics-free medium and pre-warmed to 37°C in the incubator. Cells were carefully trypsinated and counted using the Neubauer cell-counting chamber. For each replicate, $1.12x10^5$ cells were washed twice with PBS and resuspended in 8 µl buffer R. Subsequently, 2 µl Cas9-sgRNA RNPs were added, and the cells were electroporated in two pulses with a voltage of 1,000 V and a width of 40 ms. After electroporation, cells were cultured in the pre-warmed medium for at least 24 hours prior to the onset of any down-stream experiments.

2.2.2.2.3 Single cell clone generation

For the generation of *SMAD4* knockout single cell clones, PANC1 cells were electroporated in duplicates as described above and cultured for seven days. Duplicates were generated for each sgRNA (targeting exon 3, exon 6, and exon 9). After seven days, the cells were detached, pooled, and one quarter volume was taken for DNA extraction and subsequent Sanger sequencing. The remaining cells were replated in a 12-well plate (2 wells per condition) and incubated for another seven days. This was repeated two more times, each time replating the cells in a bigger culture vessel, leading to DNA from three different points in time and a 25 cm^2 cell culture flask confluent with *SMAD4* knockout PANC1 bulk cells. Additionally, one of the two wells of the 12-well plate from the first replating step was harvested, using two thirds volume for protein extraction and one third for RNA extraction.

A month after initial electroporation, the cell bulk with the sgRNA targeting exon 3 was single-cell sorted into two 96-well plates using the BioRad S3E cell sorter. The remaining cells were used for DNA extraction. After nine days, all wells with cells were detached and transferred to

a 24-well plate where they were cultured for another eleven days. Finally, cells were transferred into 12-well plates, while one sixth of them were withheld for DNA extraction. The four clones chosen for down-stream experiments were cultured further, and the other cells were frozen for long-term storage.

2.2.2.3 WST-1 cell viability assay

WST-1 assays were performed to measure the chemo-sensitivity of cell lines and single-cell clones to erlotinib and gemcitabine. Therefore, 7,500–20,000 cells were seeded in a 96-well flat bottom plate depending on their individual proliferations (Table 2.4). Cells were seeded a day before the experiment began in a 50 µl medium without antibiotics and phenol red and were allowed to attach overnight.

Table 2.4: Amount of cells used for WST-1 assays, for each of the cell lines/ single cell clones.

Cell line	Cells
HPAC	7,500
BXPC 3	7,500
CAPAN 1	10,000
CAPAN 2	20,000
PANC 1	7,500
Clone 2	7,500
Clone 17	7,500
Clone 19	7,500
Clone 20	7,500

The next day, 50 µl of medium containing serial dilutions of erlotinib, gemcitabine, or a combination of both was added. Since erlotinib was dissolved in DMSO, the highest erlotinib concentration was chosen to be a 1:200 dilution of the original concentration, limiting the final DMSO amount to 0.5%. All erlotinib dilutions were conducted in medium containing 0.5% DMSO to ensure even DMSO concentrations throughout all wells. Cells were incubated with the substances for 96 hours, after which the WST-1 proliferation reagent was added and incubated for three hours at 37°C. Absorbance was measured at 460nm in a Tecan Infinite microplate reader.

2.2.2.4 Cell-based assay with focal microscopy read-out via IncuCyte

2.2.2.4.1 Cell proliferation assay

To determine cell proliferation, $4x10^3$ cells were incubated in a 100 µl cell culture medium in a 96-well, flat-bottom plate. They were incubated for 8 days at cell culture conditions inside the IncuCyte Live-Cell Analysis Systems. Cell density was measured every eight hours using HD

phase imaging. The first picture was taken 30 minutes after the cells were placed inside the IncuCyte to allow them to settle on the bottom of the well. Border wells were filled with 200 µl PBS to reduce evaporation effects. Each condition was detected in quadruplicates.

2.2.2.4.2 Wound scratch assay

To measure cell migration, cells were seeded at 90% confluence in 1 ml within a 24-well plate and incubated overnight ($1.5x10^5$ cells/ml). The next day, the medium was removed, cells were rinsed with PBS bevor, and a scratch was applied in the middle of the well using a 10 µl pipette tip. To remove any remaining cells from the scratch, the samples were rinsed with PBS a second time, and a cell culture medium with gemcitabine, erlotinib, or a combination of both was added to the cells. When considering concentrations of gemcitabine, I chose the mean IC30 concentrations as determined by WST-1 assay (Table 2.5). For erlotinib, the highest concentration, with no expected DMSO interference (0.5% DMSO), was used.

Table 2.5: Concentration of gemcitabine and erlotinib used for wound scratch assay.

Cell line	Concentration gemcitabine [nM]	Concentration erlotinib [µM]
PANC1	53202	60
Clone 2	53202	60
Clone 17	45	60
Clone 19	45	60
Clone 20	45	60

Following the medium change, the plates were placed inside the IncuCyte and measured every three hours for four days using HD phase imaging. Border wells were filled with 200µl PBS to reduce evaporation effects, and each condition was detected in triplicates.

2.2.2.5 Knock-down of *MAPK9* via siRNA

Decreasing the *MAPK9* RNA expression in the PDAC cell lines was achieved by RNA interference (RNAi). Therefore, cells were transfected with 1–25 pmol of *MAPK9* or *GFP* siRNA (as negative control) using the Lipofectamine® RNAiMAX reagent. Depending on the experiment, cells were seeded to 60–80% confluence in 96-, 24-, or 6-well plates, always one day in advance. For the siRNA transfection, first Lipofectamine® RNAiMAX and the siRNA were each diluted in Opti-MEM® mediums 1:17 and 1:50 respectively (Table 2.6). After this, both dilutions were mixed 1:1 and incubated at RT for five minutes. The mixture was then added dropwise to the cells and incubated for 2–4 days at 37°C.

Table 2.6: Final mixture of siRNA and Lipofectamine® RNAiMAX for siRNA dependent *MAPK9* knockdown.

Component	96-well	24-well	6-well
Cells per well	$2x10^4$ cells	$5x10^4$ cells	$3.2x10^5$ cells
siRNA-lipid complex per well	10 µl	50 µl	250 µl
Final siRNA amount per well	1 pmol	5 pmol	25 pmol
Final Lipofectamine® RNAiMAX volume per well	0.3 µl	1.5 µl	7.5 µl

For *MAPK9*-dependent erlotinib sensitivity tests, erlotinib or medium was added one day after siRNA transfection and incubated together for three more days prior to the WST-1 readout. To confirm the siRNA knockdown, RNA was extracted and used for qPCR experiments after two days of incubation with siRNA. Protein extraction for the western blot analysis was performed four days after the siRNA transfection.

2.2.3 Molecular Biology

2.2.3.1 DNA and RNA isolation

2.2.3.1.1 From cell lines

For DNA or RNA isolation from cell lines between $1x10^4$ and $5x10^6$, cells were detached, centrifuged, and, after removal of cell culture medium, either directly processed or stored at -20°C. For DNA extraction, the pellet was resuspended in 200 µl PBS containing 20 µl Proteinase K, and cells were lysed by shear force in a syringe. Following this step, DNA was isolated using the Qiagen DNA Mini Kit according to the manufacturer's instructions. For RNA extraction, cells were resuspended in RLT buffer with β-mercaptoethanol, and cells were lysed using Qiashredder columns. Then, RNA was isolated with the RNeasy Mini Kit from Qiagen according to the manufacturer's instructions.

2.2.3.1.2 From patient material

Paraffin-embedded blocks from the preintervention surgical resection were assessed visually by a pathologist to determine the highest tumour content areas prior to macrodissection. DNA and RNA were extracted by ZeBanC using the half-automated Maxwell system (Promega) with the Maxwell 16 FFPE Tissue LEV DNA Purification Kit (Promega) and the RSC RNA FFPE Kit (Promega) according to the manufacturer's instructions. DNA and RNA of adequate quality were extracted from 307 and 250 FFPE samples with a tumour content of at least 10%, respectively (Supplemental Table 1).

2.2.3.2 PCR amplification from genomic DNA

Enrichment and amplification of genomic DNA was performed using the HotStarTaq DNA polymerase combined with the appropriate primer pair. Input DNA amount ranged from 0.5–100ng and final reaction volume was 25µl (Table 2.7).

Table 2.7: Composition of master mix for PCR from genomic DNA, using the HotStarTaq DNA polymerase.

Component	Volume
HotStar Taq polymerase	0.125 µl
10x PCR buffer	2.5 µl
dNTP mix (25 mM each)	0.2 µl
Forward primer (10 µM)	1 µl
Reverse primer (10 µM)	1 µl
Genomic DNA	0.5-100 ng
H_2O	Ad 25 µl

Cycling conditions were adjusted if necessary but generally followed the parameters advised by the manufacturer (Table 2.8).

Table 2.8: HotStar Taq DNA Polymerase PCR program.

Step	Temperature	Time	Cycles
Initial denaturation	95 °C	15 min	
Denaturation	94 °C	20 s	30x
Annealing	50-65 °C	20 s	
Extension	72 °C	1 min	
Final extension	72 °C	7 min	
Hold	15 °C	∞	

After amplification, 2 µl PCR products were mixed with 6x loading dye and 10x gel red before being separated by size on a 1.5% agarose gel in a TAE buffer for 30–60 minutes at 120 V. The agarose gels were imaged with UV light on an AlphaImager. If necessary for downstream experiments, amplification products were purified using AMPure Beads XP or a Qiagen PCR purification kit according to the manufacturer's instructions.

2.2.3.3 Digital droplet PCR (ddPCR)

KRAS G12/G13 variants were validated using ddPCR. This form of digital PCR is based on a single, real-time PCR reaction with fluorescence-based probes that takes place within small oil droplets and leads to a high-sensitivity variant calling of up to 0.5% variant allele frequency

(VAF). For ddPCR, all reagents of the ddPCR™ *KRAS* G12/G13 Screening Kit were mixed with 20ng patient DNA according to the manufacturer's instructions (Table 2.9).

Table 2.9: Master mix for ddPCR.

Component	Volume/amount
2x ddPCR Supermix for Probes (no dUTP)	10 µl
20x multiplex primers/probes (FAM + HEX)	1 µl
Genomic DNA	20 ng
H_2O	ad 20 µl

Including a positive and a negative control, 20µl master mix per sample was poured into a DG8™ Cartridge together with a 70µl droplet of generation oil. Droplets were generated in a QX200 droplet generator according to the manufacturer's instruction. The droplets were carefully transferred into a clean, 96-well plate, and the plate was sealed. Following this, cycling conditions were used to amplify the genomic DNA according to the manufacturer's instructions (Table 2.10).

Table 2.10: Cycling conditions for amplification of ddPCR reaction.

Step	Temperature	Time	Ramp rate	Cycles
Enzyme activation	95 °C	10 min		
Denaturation	94 °C	30 s	2 °C/s	40x
Annealing/extension	55 °C	1 min		
Enzyme deactivation	98 °C	10 min		
Hold	4 °C	∞	1 °C/s	

Then, the plate was placed in a QX200 Droplet reader for droplet quantification. Data acquisition and analysis were performed using QuantaLife software. The software calculates DNA concentration in copies per µl. Variant allele fraction can be calculated by dividing the droplet concentrations of target (FAM channel) and reference (HEX channel).

2.2.3.4 Reverse transcription

To transcribe RNA into cDNA, the Invitrogen™ M-MLV reverse transcriptase was used according to the manufacturer's instructions. Random primers and RNA were mixed with dNTPs (Table 2.11) and heated to 65 °C for 5 minutes and quick chilled on ice.

Table 2.11: Master mix for first step of reverse transcription using M-MLV reverse transcriptase.

Component	Volume
250 ng random primers	1 µl
Total RNA	200 ng – 1 µg
10 mM dNTP Mix	1 µl
Sterile distilled H_2O	Ad 12µl

Next, buffer and DTT were added (Table 2.12) and incubated for two minutes at 37°C.

Table 2.12: Master mix for second step of reverse transcription using M-MLV reverse transcriptase.

Component	Volume
5x First-Strand Buffer	4 µl
0.1 M DTT	2 µl
H_2O	1 µl

Finally, a 1µl M-MLV reverse transcriptase was added and incubated for 10 minutes at 25°C, followed by 50 min at 37°C and a heat inactivation of the transcriptase at 70°C for 15 minutes. If necessary, cDNA concentration was measured with a NanoDrop1000 spectrophotometer.

2.2.3.5 Quantitative real-time PCR (qPCR)

For RNA quantification, the Luna® universal qPCR master mix was used according to manufacturer's instructions. All components were mixed (Table 2.13) and cycled on an Applied Biosystems StepOnePlus® under the recommended cycling conditions (Table 2.14). Data acquisition and analysis were performed with the StepOne Software, and relative target expression was calculated via the ΔΔCT-method.

Table 2.13: Master mix for qPCR with the Luna® universal qPCR Kit.

Component	20 µl reaction	Final concentration
Luna Universal qPCR Master Mix	10 µl	1X
Forward primer (10 µM)	0.5 µl	0.25 µM
Reverse primer (10 µM)	0.5 µl	0.25 µM
Template cDNA	variable	< 100 ng
Nuclease-free Water	to 20 µl	

Table 2.14: Light cycler program for qPCR.

Step	Temperature	Time	Cycles
Initial denaturation	95 °C	60 s	
Denaturation	95 °C	15 s	40x
Extension	55-60 °C	30 s	
Melt Curve	60-95 °C	various	

2.2.3.6 Multiplex ligation-dependent probe amplification

To validate potential CNAs identified by targeted sequencing and to identify CNAs in samples not suitable for CNA detection by targeted sequencing, we used commercially available multiplex, ligation-dependent probe amplification (MLPA) assays according to the manufacturer's instructions. Therefore, 70ng of DNA was denatured, hybridized to probes of either the SALSA MLPA probe mix P294-C1 Tumour Loss or the P175-B1 Tumour Gain (MRC Holland; Table 2.15), ligated, and amplified (Table 2.16). Each probe resulted in an amplicon with a distinct length.

Table 2.15: Composition of MLPA reaction mix.

Component	Volume
DNA	5 µl
Hybridisation master mix	3 µl
Ligase-65 master mix	32 µl
Polymerase master mix	10 µl

Table 2.16: Thermocycler program for the MLPA reaction.

Step	Temperature	Time	Cycles
DNA denaturation	98 °C	5 min	
	25 °C	pause	
Hybridisation reaction	95 °C	1 min	
	60 °C	16-20 h	
Ligation reaction	54 °C	pause	
	54 °C	15 min	
	98 °C	5 min	
	20 °C	pause	
PCR reaction	95 °C	30 sec	35x
	60 °C	30 sec	
	72 °C	1 min	
	72 °C	20 min	
	15 °C	∞	

Subsequently, the amplicon mix was separated with capillary electrophoresis, and the ratio of target to backbone probe was calculated. Probes with a ratio below 0.8 were considered deleted and above 1.2 were considered amplified. Data analysis was performed using Coffalyser.Net (version 140721.1958).

2.2.3.7 Targeted expression analysis using nCounter Technologies

For RNA expression analysis, we used the nCounter Technologies System, a targeted hybridisation-based assay with gene-specific, fluorescence-coupled probes that bind to the immobilized RNA and enable direct digital detection of expression levels. Therefore, RNA concentration was measured on a Quantus Fluorometer using the QuantiFluor RNA System (Promega), and RNA degradation levels were determined with the RNA 6000 Nano Kit on a 2100 Bioanalyzer (Agilent). To adjust for high degradation levels in FFPE RNA, sample concentration was calculated only for fragments suitable for nCounter analysis (50 nt–300 nt; size fragment) using the following equation as advised by NanoString technologies.

$$conc_{adj} = \frac{300}{300 - size\ fragment\ [\%]} * 100$$

The NanoString nCounter Flex system was used to run a customised version of the PanCancer Pathways Panel (770 genes representing 13 canonical pathways in cancer, 606 pathway genes, 124 cancer driver genes, and 40 reference genes panels) with 28 additional genes (Supplemental Table S4). Raw NanoString counts were normalised to internal positive control probes, and housekeeping genes using nSolver Software (NanoString Technologies, WA, USA) version 4.0 were, according to default parameters with a background threshold count value, set to 20. After quality controls, 230 samples were analysed. Differential expression analysis was conducted using the nCounter Advanced Analysis Module (version 2.0.115). Genes were tested for differential expression in response to each selected covariate. For each gene, a single linear regression was fit using all selected covariates to predict expression, and the false discovery rate (FDR) was estimated according to the Benjamini-Hochberg procedure[145].

2.2.4 Protein Biochemistry

2.2.4.1 Protein extraction

For protein extraction, cells were harvested and washed once with PBS before being resuspended in an appropriate amount of lysis buffer containing a protease inhibitor (cOmplete™ Roche) and a phosphatase inhibitor (PhosSTOP™, Roche). Cells were kept on ice at all times and lysed at 4°C using a Bioruptor Plus with the following settings (Table 2.17):

Table 2.17: Settings for cell lysis on Bioruptur Plus.

Condition	Setting
Cycle Number	12
Time ON	30 sec
Time OFF	30 sec

Cell fragments were subsequently removed by centrifugation for 15 minutes at 17,000g and 4°C. The supernatant was collected and stored at -20°C until further use.

2.2.4.2 BCA protein concentration measurement

Protein concentration was measured using the Pierce BCA Protein Assay Kit according to the manufacturer's instruction, though with some minor changes. For each sample, a 150 µl working reagent was mixed with a 10 µl protein solution. All samples were diluted 1:10 before being added to the mix and measuring both standard and samples in duplicates. After incubation for 30 min at 37°C, absorption was measured in a Tecan Infinite microplate reader at 562 nm.

2.2.4.3 Protein separation via SDS PAGE

For protein separation, the protein solutions were thawed on ice, and a 20–50 µg protein was mixed with a Laemmli sample buffer containing 10% β-mercaptoethanol. Protein denaturation was achieved through a five-minute incubation at 95°C followed by two minutes on ice. Then, samples were loaded onto a 10% SDS gel (Table 2.18). Proteins were subsequently separated for 1.5–3 hours at increasing voltages (starting at 60 V up to 140 V) until the desired level of separation was reached.

Table 2.18: Composition of resolving gel and stacking gel used for protein separation. The volumes are sufficient for two gels.

Reagent	Resolving gel	Stacking gel
H_20	2.8 ml	2 ml
Resolving/stacking gel buffer	2 ml	1.3 ml
Acrylamid (40 %)	2.7 ml	0.7 ml
Glycerin (40 %)	4 ml	1 ml
SDS (10 %)	50 µl	25 µl
APS (10 %)	50 µl	40 µl
TEMED	10 µl	10 µl

2.2.4.4 Western blot

Western blot analyses were performed to quantify the number of specific proteins within cell line samples. Therefore, a nitrocellulose membrane (Amersham™ Protran™ Premium 0.45 µm NC) was first moistened in Millipore H_2O for five minutes, and both the SDS PAGE as well as a nitrocellulose membrane were subsequently equilibrated in a transfer buffer for 15 minutes at RT. Next, the wet blotting chamber was assembled as seen in Figure 2.2. and run at 320 mA for 70 minutes or at 120 mA overnight at 4°C.

Figure 2.2: Western blot set-up

After blotting, the membrane was dyed with acidic red ponceau solution and cut depending on protein size. Membranes were then blocked with 5% BSA in TBS-T for at least 60 minutes at RT, followed by incubation with primary antibodies overnight at 4° C. The next day, membranes were washed 3x for five minutes with TBS-T at RT and subsequently incubated with horseradish peroxidase, labelled secondary antibody, for at least 60 minutes. After this, the membrane was washed 3x for 10 min with TBS-T at RT prior to detection. Both primary and secondary antibodies were diluted according to manufacturer instructions. For protein detection, both components of the ECL™ prime western blotting detection reagents were mixed, and an appropriate amount was pipetted onto the membranes. The subsequent chemiluminescence reaction was detected with an ImageQuant™ LAS 4000.

2.2.5 Sequencing Methods

2.2.5.1 Sanger sequencing

Validation of SNVs in cell lines was usually conducted via Sanger sequencing. To accomplish this, the region of interest was first amplified by PCR. The PCR product was then purified using either the QIAquick PCR Purification Kit or AMPure XP beads (Agencourt) according to the manufacturer's instructions. The concentration of purified PCR amplicons was measured using a Nanotrop1000 Spectrometer. Following manufacturer instructions, amplicons were subsequently diluted to 10–40ng/µl, and 5µl DNA was mixed with 5µl forward or reverse primer. The mixture was sent to Eurofins Genomics for sequencing. Resulting sequences were analysed with Mutation Surveyor Software (version 5.0.0) and for CRISPR-knockout clones; the proportion of the different genetic events was calculated using the Synthego ICE analysis tool[146].

2.2.5.2 Amplicon based deep sequencing

For amplicon-based targeted deep sequencing, PCR-amplicons of approximately 100bp–200bp length were generated from genomic DNA covering the area of the expected variant. Amplicons were subsequently pooled for library construction and indexed using the NEBNext Ultra DNA Library Prep Kit (New England Biolabs). The pools were paired-end sequenced on a MiSeq sequencer using the MiSeq Reagent Kit v2 (300 cycle, Illumina).

2.2.5.3 Targeted next generation sequencing

2.2.5.3.1 Panel generation and sample quality control

The PDAC specific custom next generation sequencing (NGS) panel covered full-length coding regions of 67 genes described as 'significantly mutated' in one of the two largest PDAC sequencing studies[101,102] or were shown as clinically relevant in PDAC[147-149]. I also added genes that were previously included in PDAC gene panels[148,150], and/or representing major players in the EGFR pathway (Supplemental Table 2). To test its specificity and overall coverage, 16 patient samples with the highest DNA concentrations were sequenced in an initial test run. The results were also used to adapt library preparation conditions, number of samples per HiSeq cartridge, and error detection during variant calling. As FFPE DNA is known for its poor quality, all samples were tested for their amplifiability by running a WT1 single nucleotide polymorphism (SNP) TaqMan assay prior to library generation. Subsequently, sample concentration was determined with the Qbit fluorescence system, and sample fragmentation was measured on the 4200 TapeStation System using Genomic DNA ScreenTapes (Agilent).

2.2.5.3.2 Library Prep and Sequencing

The custom Agilent SureSelectXT Target Enrichment System for Illumina Paired-End Multiplexed Sequencing was used according to the manufacturer's instructions (with some adaptations) to facilitate library preparation of DNA from patient samples. Library preparation was begun with 200ng DNA in a 50µl low TE buffer. Samples were fragmented with an instrument of Covaris model E220 using 130µl Covaris microTUBEs to target DNA fragment sizes of 150 to 200 bp. Both shearing conditions and PCR cycles were adjusted according to each sample's genomic DNA peak size during library amplification (Table 2.19). To adjust for the poor quality of FFPE samples, the KAPA Hyper Prep Kit (Kapa Biosystems) was used for end repair and A-tailing per manufacturer instructions. After amplification of the adapter-ligated libraries, a quality check was performed using an Agilent 4200 TapeStation and a D1000 ScreenTape. A 350ng–750ng DNA library with a peak size between 200bp and 300 bp was included in the hybridization and capture step. Samples with lower DNA concentrations could be re-amplified by up to three cycles to improve yield. If that was not enough and sufficient patient material was available, library generation was repeated with a 400ng starting DNA. If

not, samples were considered failed during library prep and were excluded from subsequent analysis.

Table 2.19: Shearing conditions and PCR cycles for adapter library amplification, adjusted for DNA Fragmentation.

Peak size genomic DNA [bp]	Shearing time [s]	PCR Cycles
< 420	90	10
420 - 1200	120	11
> 1200	150	12

For sequencing, 48 samples were pooled equimolar in a 2 nM sample mix and paired-end sequenced with a mean sequencing depth of ~600x and a minimal reading depth of 200x on a HiSeq 2000 sequencer (Illumina).

For cell line DNA, library preparation was conducted as described above but with added use of the KAPA Hyper Prep Kit. The higher DNA quality allowed for the Agilent SureSelectXT Target Enrichment System to be sued for end repair and A-tailing.

2.2.5.3.3 SNV Calling

Sequence alignment and initial mutation calling were performed by our Japanese cooperation partners from the lab of Professor Seishi Ogawa using their in-house pipeline 'Genomon v.2.5.0', as previously described[151]. Readings showing either a mapping quality score of <25, a base quality score of <30, or 5 or more mismatched bases were excluded from the analysis. Candidate mutations with i) a variant allele frequency (VAF) ≥0.04 and ii) an EBcall[152] (Empirical Bayesian mutation calling) p-value $\leq 1\times10^{-10}$ were adopted and filtered further. The filtering was conducted by using the following criteria: I excluded i) the synonymous mutations and variants outside the targeted areas of the 67-gene panel; ii) known variants listed in the 1000 Genomes Project (version May 2011) and the NCBI SNP database (dbSNP) build 131 as well as data from the National Heart, Lung, and Blood Institute (NHLBI) Exome Sequencing Project (ESP) 5400, and information from the Human Genome Variation Database (HGVD; October 2013 release); iii) variants present only in unidirectional reads; iv) variants occurring in repetitive genomic regions; v) variants with <2 supporting reads; vi) all variants found in non-paired normal samples (n=20) showing an allele frequency of >0.0025; vii) samples with a VAF of <4% or between 40–60%; and viii) EBcall p-value <4. Samples with VAF outside the cut-off values or a low EBcall p-value were rescued if they were mentioned in the COSMIC database at least 10 times in cancer-related areas. Finally, mapping errors and potential germline

variants were removed by visual inspection in the Integrative Genomics Viewer (IGV) browser[153].

2.2.5.3.4 CNA Calling

Copy-number analysis was performed as previously reported[154]. This was accomplished using an in-house pipeline (Y. Shiozawa and S. Ogawa, manuscript in preparation) in input data comprised the total number of reads covering each bait region and the allele frequency of heterozygous single-nucleotide polymorphisms (SNP) detected by targeted sequencing. To this aim, 1,305 probes spread across the entire genome to serve as a chromosomal backbone were included in the custom panel. Along with 400 gene-specific probes (3–6 per gene) targeting 100 genes previously described as drivers of PDAC (

Supplemental Table 3), this enabled CNA detection. The VAF of each SNP was calculated and used to determine a baseline with a total copy number (tCN) of 2 and an allelic ration (AR; ratio of maternal/paternal allele) of 1. While regions deviating from this were detected by the pipeline and flagged as either deletion (tCN<2, AR<1), amplification (tCN>2, AR<1), or uniparental disomy (UPD) (tCN=2, AR<1), each variant had to be inspected manually to account for bad sample quality and low tumour content. After careful evaluation, I chose the initial cut-off values of tCN≤1.75 and AR≤0.75 for deletions and tCN≥2.25 and AR≤0.75 for amplification, but individual variants were excluded or included independent of the cut-off after manual inspection.

2.2.6 Statistical analysis

Statistical analyses were performed using IBM SPSS statistics (version 24) and R version 3.6.1. Co-occurrence and mutational exclusivity were calculated with Fisher's exact test and subsequently corrected for multiple hypothesis testing using the Benjamini–Hochberg method[145]. Non-negative matrix factorization, which is an unsupervised machine-learning approach, was performed using the R-package *Bratwurst* (version 1.0)[155] to identify metagenes based on SNV and CNA patterns of investigated patients. I used the mutation status of all 67 panel genes as well as the CNA status of the 11 genes included with the MLPA tumour loss kit as input. The optimal number of metagenes (k=4) was selected based on several quality criteria (e.g., Frobenius reconstruction error, cophenetic correlation coefficient, and amari type distance). Based on their metages, patients were clustered hierarchically to define each specific cluster. For alteration enrichment analysis in the NMF clusters, I calculated odds ratios (OR) for each alteration and compared them with two-tailed t-tests. Variants were considered significantly enriched when, after multiple testing corrections, p-value was <0.05. The Fisher's exact test was subsequently used to compare the number of alterations and base-change pattern. Cox models were used for time-to-event variables (OS and DFS), and p-values were calculated using the Wald test. Multivariate cox proportional hazards models were used to investigate variables associated with survival endpoints. Missing CA19-9 values (approximately 10% of patients) were filled using multiple imputation with 10x imputations. To select input for the multivariate cox proportional hazards models, univariate cox regression analysis of all clinical variables and all genetic alterations were carried out and then adjusted for multiple testing using the Benjamin-Hochberg procedure[145]. Primary analysis endpoint was OS, followed by exploratory analysis into DFS. While genetic alterations were encoded as either 'altered' or 'normal', clinical variables were dichotomized as follows: tumour size (T1/2 vs T3/4), lymph node involvement (N0 vs N+), grade (grade 1 to 2 vs grade 3), Karnofsky performance status (90% to 100% vs < 90%), CA19-9 (≤100 vs >100 kU/L), age (≤65 vs >65), sex (m vs f), and treatment arm (gemcitabine vs gemcitabine+erlotinib). For gene expression values, dichotomisation was achieved by splitting patients based on each gene's normalised

median expression. Variables with an adjusted p-value ≤ 0.1 were included in subsequent multivariate analyses. To determine whether the effect of adding erlotinib to the gemcitabine standard therapy differed between distinct patient subgroups, a test of interaction was conducted based on the cox proportional hazard regression model – a model which collectively included all patients in a single analysis-evaluated treatment arm and patient subgroup while including relevant clinical variables status as a three-level variable. A model with an interaction term between patient subgroups and treatment arm allowed different treatment based on relative risks for patients in the specific subgroup and the rest of the CO-2016 cohort. Its log partial likelihood was compared with that of a model without interaction to test whether the treatment of relative risks differed significantly by receptor status. The log partial likelihood ratio test was used[156]. For the maximally selected rank statistics with several p-value approximations (maxstat), we used the R-package *maxstat* (Version 0.7-25) on all genes with a median expression of ≥200 and a standard deviation ≥ 100. Results were adjusted for multiple testing with the Benjamini–Hochberg method[145]. A Kaplan-Meier analysis was performed to construct survival curves, and a log-rank test was applied to evaluate differences among subgroups.

3 Results

3.1 Targeted NGS with PDAC-specific custom panel

The genetic background of pancreatic cancer has been analysed in several large sequencing studies. I could use this existing knowledge to focus on a number of genes that were previously shown to be regularly mutated in PDAC, including genes theorized to affect clinical outcome in general and erlotinib sensitivity in particular. To this end, I designed a custom next generation sequencing (NGS) panel that covered the full-length coding region of 67 genes as well as gene specific probes facilitating the CNA detection in another 67 partly overlapping genes. Quality control of the panel and the sequenced samples as well as strict variant calling criteria and validation with a second independent method ensured a high quality data set, which served as a basis for the subsequent biomarker search.

3.1.1 Panel design and quality control

To select the genes to include in the custom NGS PDAC panel, I extracted the most relevant targets using data from previous whole exome/genome sequencing studies and published data about potential prognostic and predictive biomarkers in PDAC. The final panel included 100 genes with known significance in PDAC and EGFR signalling (Supplemental Table 2,

Supplemental Table 3).

The basis of my NGS panel design for SNV detection was a complex scoring system I developed to evaluate the impact each gene had on three important characteristics and thereby to detect the most important PDAC genes. The first parameter was how frequently the gene was found to be mutated in one of two major whole-exome/whole-genome sequencing studies that analysed a total of 556 PDAC samples[101,102] (Figure 3.1 'Mutation frequency in literature'). All 9,934 genes from these papers were categorized into one of six groups and scored accordingly (Table 3.1). Each gene was scored for each of the two papers, making 10 the highest possible mutation score.

Table 3.1: Mutation score for SNV genes in PDAC specific targeted NGS panel design.

Mutation frequency	Score
50 % - 100 %	5
20 % - 49 %	4
10 % - 19 %	3
3 % - 9 %	2
1 % - 2.9 %	1
<1 %	0.5

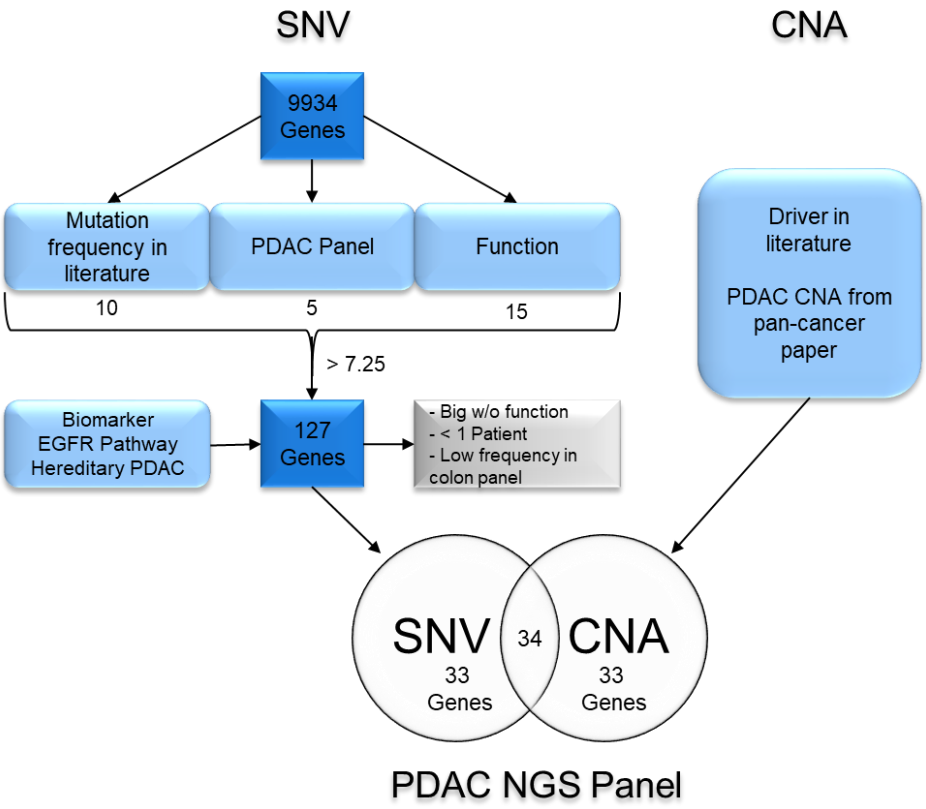


Figure 3.1: Flowchart of the custom PDAC next generation sequencing panel design. To identify relevant genes for SNVs, all 9,934 genes from two whole-exome/genome sequencing studies[101,102] were scored according to three parameters and subsequently filtered based on their scoring sum and additional independent parameters (grey box). Genes for CNA detection were collected from different sources and combined with the SNV genes into the final PDAC NGS panel. The numbers below the NGS input parameter indicate the highest score for each parameter in the scoring system. The number next to the arrow indicates the cut-off value for the scoring sum which was necessary for each gene to be selected for the first panel draft with 127 genes.

The second parameter on which panel inclusion was decided was whether they had been selected in previous PDAC gene panels (Figure 3.1 'PDAC Panel'). To this end, two panels used for targeted NGS on PDAC in an experimental setting[148,157] and three commercial panels[150,158,159], used predominantly to test patients with familiar predispositions, were sighted. Each of the 9,934 genes from the two major whole-exome/whole-genome sequencing studies[101,102] was scored one point for each panel in which it had been included, making five the highest possible score for this category. The last parameter was a functional score (Figure 3.1 'Function'). Five points were scored to each gene if it was found in one of the following three categories: (i) driver mutations according to Waddel et al.[101] or Bailey et al.[102], (ii) part of the EGFR or closely related pathways (mTOR, PIK-/Akt, MAPK, NFκB)[160], or (iii) clinically relevant or biomarkers in PDAC[147-149,161]. Finally, the previous three parameters were combined to calculate a summary score for each gene, with a maximum possible score of 30. I included all genes with a score of 7.25 or higher. Since the aim of the subsequent analysis was to find potential biomarker for erlotinib sensitivity, I wanted to prevent possible candidates

within the EGFR pathway or known clinical associations to be lost because of low mutation frequency. Therefore, I also included genes which did not have a score above 7.25: five genes with important functional consequence, 38 genes from the EGFR pathway, and eight genes known to be predictors in familiar PDAC. This first panel was comprised of 127 genes and needed to be decreased further to fit the size requirements feasible for our targeted NGS. Therefore, I excluded especially large genes (<10 kb) with no known statuses as PDAC drivers and without known functional or clinical consequences for PDAC. Next, I used the mutation frequencies from Bailey et al.[102] to extrapolate any genes which would be detected in less than one patient in the slightly smaller CO-2016 cohort. This filtering led to the final panel of 67 genes, which could be used for SNV detection (Supplemental Table 2). The gene-specific probes for CNA detection were then chosen based on intensive literature research. This included the two major whole-exome/whole-genome studies[101,102] as well as several publications about the landscape of somatic CNAs across human cancers[157,162-165]. Combination of these sources identified 67 genes that had been frequently altered in PDAC, 34 of which overlapped with genes included for SNV detection (Figure 3.1). The final panel included probes covering full-length exomes of 67 SNV genes, probes covering SNPs in and around 67 genes for CNA detection, and 1,305 probes against SNPs covering the entire chromosome backbone. The covered region had a size of 0.7MB and was designed using the SureDesign Platform (Agilent) in cooperation with Agilent design experts.

To ensure an even coverage of all areas included in the custom panel, I began by sequencing nine patient samples and analysing the mean coverage for each gene. The genes were covered equally with a mean coverage between 500 and 1,250 reads (Figure 3.2 A). Within the genes, the coverage was largely homogenous, with only a few outliers either under- or overrepresented (Figure 3.2 B, C). In summary, only three exons were below the cut-off value of 200 reads; therefore, the newly designed PDAC panel could be used for high-quality sequencing of all 67 included genes.

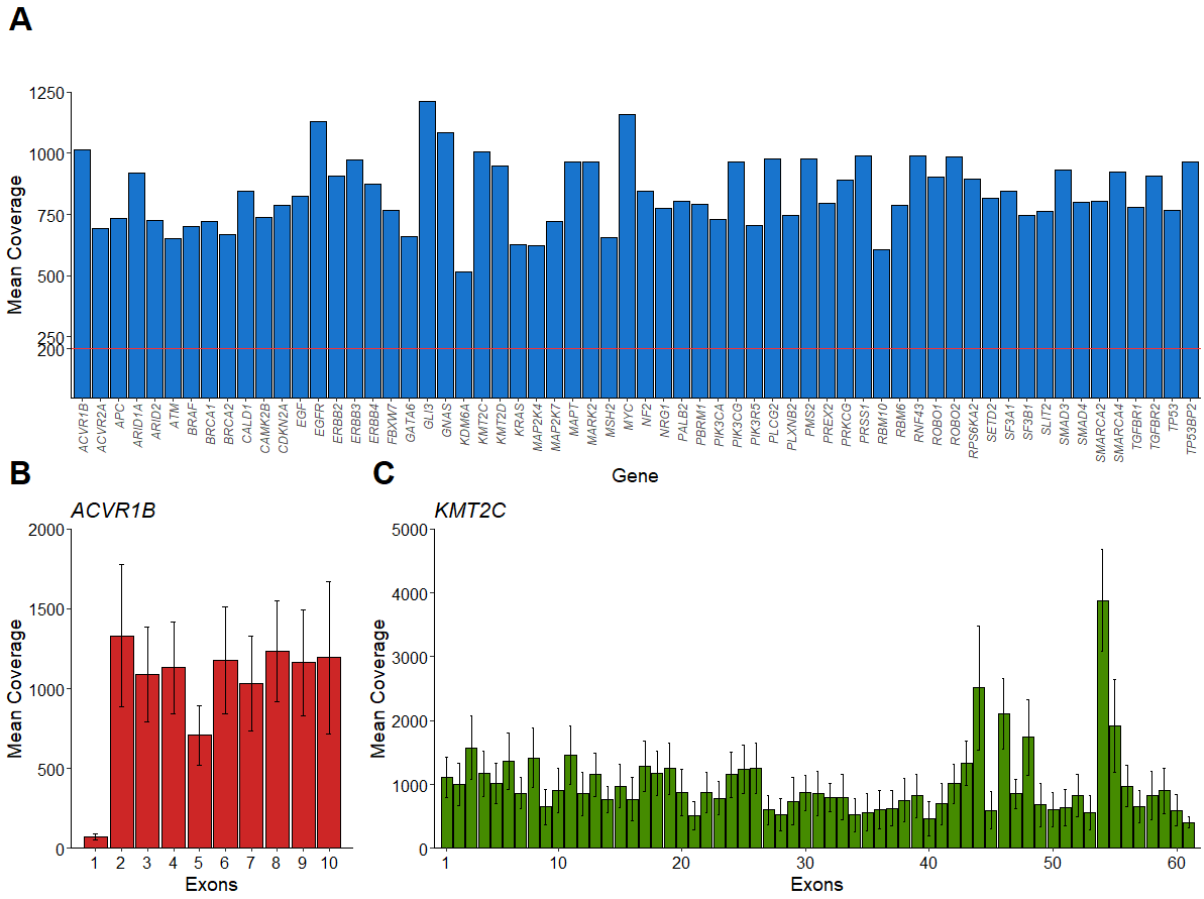


Figure 3.2: Overview of the coverage of all 67 genes included in the custom panel. Mean Coverage was calculated for nine patient samples**.** The bars ± error bars show mean ± standard deviation. **A)** Mean Coverage of all exons in each of the 67 genes showed an even read distribution and a mean coverage of all genes above the 200 reads threshold (red line). **B)** An example of a gene with an underrepresented region is *ACVR1.* It is one of only three genes with an exon that is covered with fewer than 200 reads. **C)** An example of a gene with an overrepresented region, *KMT2C,* is generally very well-covered, with five exons showing a coverage more than three times higher than the panel-wide average.

3.1.2 Sample quality control

Since FPPE DNA is known to be highly degraded, a quality control step was necessary before commencing sequencing. First, all samples were included in a *WT1* SNP assay to assess their amplifiability. All samples could be amplified sufficiently for *WT1* status determination. The *WT1* status distribution did not differ significantly from those known for healthy individuals (homozygous C/C =78%, heterozygous C/A =21%, homozygous A/A =1%). Nevertheless, comparison of amplification curves against curves generated using DNA from fresh patient blood revealed a wider spread in ΔCT values within the FFPE sample cohort and a trend toward delayed amplification (Figure 3.3 A). To quantify differences in DNA degradation and to adjust library preparation for each sample accordingly, gDNA from all samples was

measured using an Agilent TapeStation. DNA peak size differed widely from below 500 bp to above 3500 bp (Figure 3.3 B).

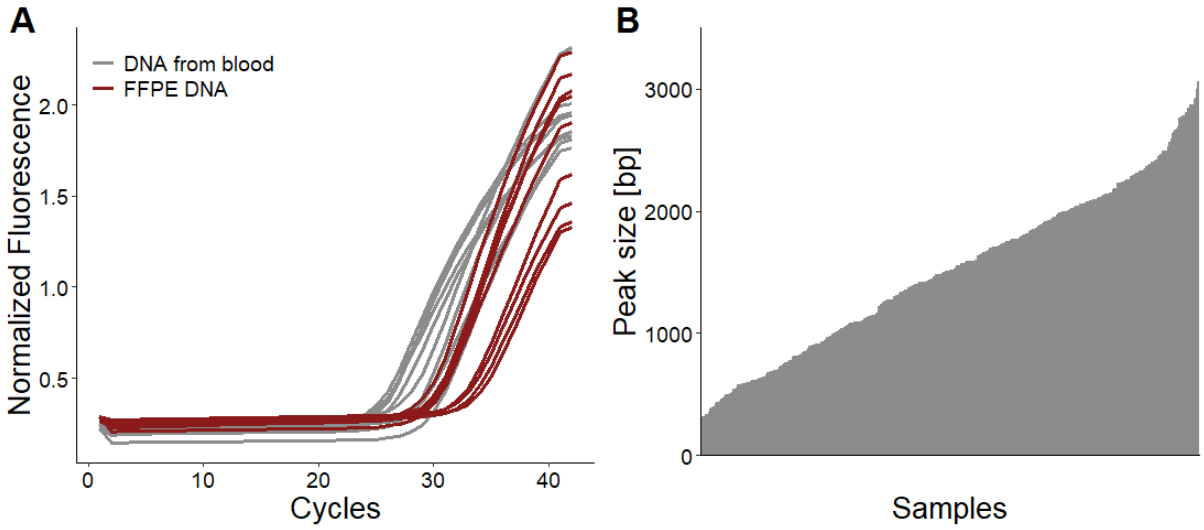


Figure 3.3: Quality control criteria for FFPE patient samples. A) Normalized fluorescence from qPCR of nine exemplary patient samples compared to nine samples extracted from fresh patient blood. **B)** Peak size of genomic DNA from 340 sequenced samples.

Despite interspersed DNA quality, all but two samples could be sequenced successfully. After completing sequencing, the quality of the final data set was assessed based on four criteria: (i) average sequencing depth of each sample, (ii) ratio of amplicons with a sequencing depth of at least 100 (ratio 100x), (iii) ratio of amplicons that are PCR duplicates (duplication rate), and (iv) ratio of number of bases within the target region to total number of bases (on-target ratio). The samples had a mean reading depth of 606 ± 225 (mean ± SD). I chose a mean reading depth of 200 reads per sample as the cut-off value, excluding five samples from subsequent analysis (Figure 3.4 A). With a mean 100x ratio of 98%, the samples were evenly covered even after removal of the relatively high number of duplicate reads. High fragmentation of gDNA necessitated an increased number of PCR cycles during library preparation, which led to comparatively high duplication rates. Since duplicate reads were excluded from the final analysis, the duplication rate directly impacts each sample's reading depth (Figure 3.4 B and C). While there was a direct correlation among all four characteristics, the correlation with gDNA peak size, and thereby DNA quality, was only marginal (Figure 3.4, Supplemental Figure 1 and Supplemental Figure 2). Sequencing quality varied by sample. Despite this, all patients (excluding 5, which did not pass quality control) could be sequenced, generating high-quality sequencing data for 300 FFPE PDAC samples.

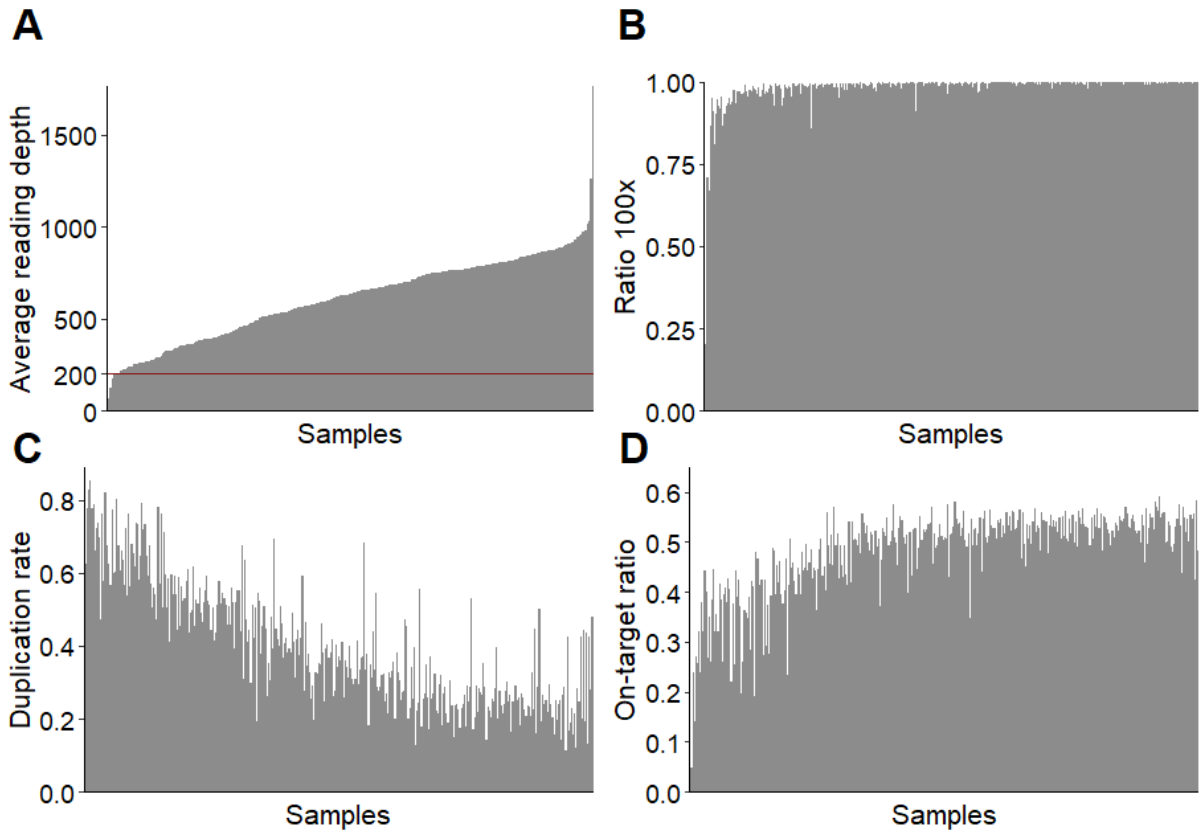


Figure 3.4: Quality control criteria for sequenced samples (n=307). All samples are ordered according to increasing average reading depth. **A)** Average reading depth of sequenced cohort after exclusion of duplicate reads. The cut-off value of 200 reads for inclusion in subsequent analysis is marked with a red line. Only five samples had a reading depth below 200. **B)** Ratio of amplicons with a reading depth of at least 100 shows even coverage throughout the cohort. **C)** Ratio of PCR duplicates within each sample. **D)** Ratio of reads that map to the area included in the targeted gene panel.

3.1.3 Variant calling and validation

Of the 436 patients enrolled in the Conko-005 study, we collected tumour tissue from 331 patients. Due to low tumour content, 24 tissues needed to be excluded, leaving 307 for sequencing. Two samples failed during library preparation, and five samples were subsequently excluded because of low average reading depth (< 200x). Seven additional samples had to be excluded because of missing clinical data, leaving 293 patients for final analysis (CO-2016 cohort). Of those, 10 patients could not be analysed for copy number alterations, neither with NGS nor with MLPA. As a result, I then had information on SNV and CNA status for all 11 genes in the MLPA panel for 283 patients. For approximately two thirds of these patients (n=171), CNA analysis based on NGS and MLPA was successful, which led to SNV and CNA status for all 100 genes included in the panel (Figure 3.5).

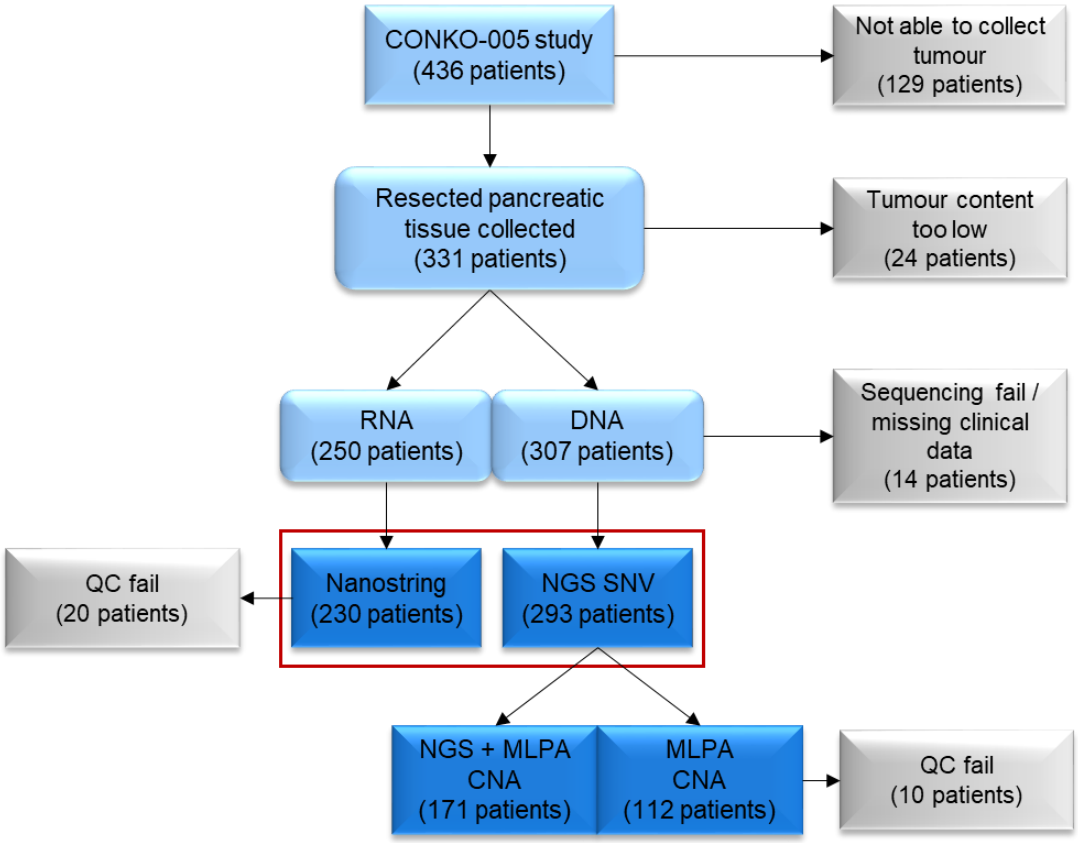


Figure 3.5: Flow chart of sample procession. Shows the number of samples included in each analysis for each step, as well the amount of and reason for sample exclusion. The red box marks the 293 sample comprising CO-2016 cohort, which is the basis of most subsequent analysis. Modified from Hoyer et al.[1]

3.1.3.1 Single nucleotide variants in R0-resected PDAC

To reduce the likelihood of false positive SNV calling, I established a validation pipeline using a combination of several variables such as SNV frequency, EB call p-value, DNA quality, and sequencing duplication rate. In a first validation run, I screened variants with a wide range of these quality criteria representing the whole spectrum of variants in the CO-2016 cohort. This initial run had a validation rate of 82%. Closer analysis showed that the relatively low validation rate was due to 'low quality variants' (Low) with specific characteristics. Therefore, in a second validation run, I targeted these low-quality variants by validating all samples with an EB-call <12, a VAF <10, and a total sample reading depth <400. Additionally, I validated all *SMAD4* and *ATM* variants. In total, 256 potential SNVs representing 19% of all detected variants were investigated in a second independent experiment either through amplicon-based targeted deep sequencing (n=231) or ddPCR (n=24) as previously described[166]. This led to a validation rate of 95% for good-quality variants (representing 94% of all variants) and 64% for low-quality variants (representing 6% of all variants). Of note, a subset of *KRAS* G12/G13 variants was chosen for validation with ddPCR and could be validated completely (Table 3.2). Following

removal of all false positive, low-quality variants, the false positive rate in the final variant list was <5 %.

Table 3.2: Overview of the results from both validation runs and the ddPCR validation of a selection of *KRAS* G12/G13 variants. Low: low-quality filter criteria: EB-call <12, VAF <10 and total sample reading depth <400 .

	Run1	Run2	ddPCR	Low	Others
Variants	65	183	24	81	192
PCR failed	4	13	0	7	10
Validated	50	144	24	47	172
Not Validated	11	26	0	27	10
Validation rate	**82 %**	**85 %**	**100 %**	**64 %**	**95 %**

For the validated variant list, the number of mutations as well as their respective allele burdens were independent from morphological estimation of tumour content, as shown by the extremely low R^2 values (Figure 3.6).

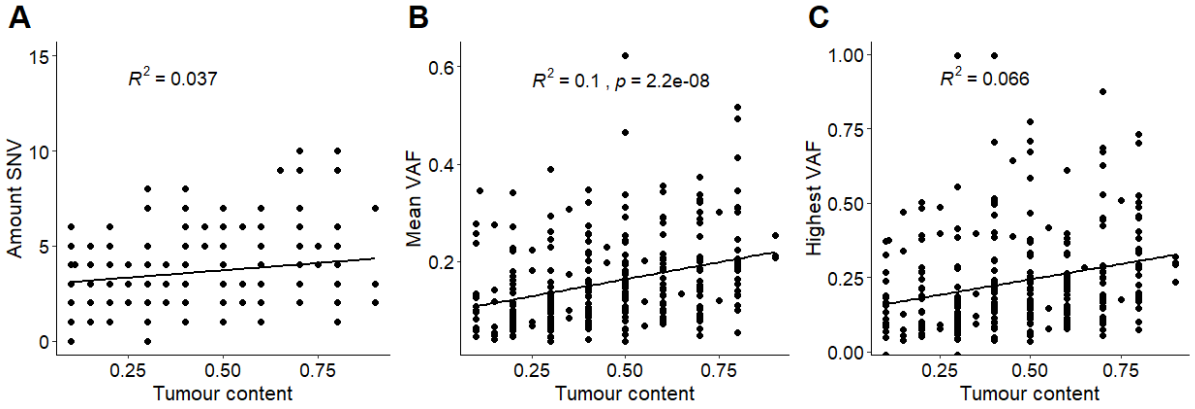


Figure 3.6: Comparison of tumour content with SNV characteristics. No correlation was found between morphological estimation of tumour content and **A)** amount of SNV per patient, **B)** mean VAF of all variants in one patient, or **C)** highest VAF. This is exemplified by the low proportion of variation in the dependent (response) variable (R^2), which is in turn explained by the linear regression model. Modified from Hoyer et al.[1]

A total of 1086 SNVs were identified in 293 patients (Supplemental Table 4). Most SNVs were missense mutations (69%) followed by frameshift insertions or deletions (indels, 14%) and stop gains (11%). Within the 67 genes analysed for SNVs, 58 were found to be recurrently mutated (Figure 3.7).

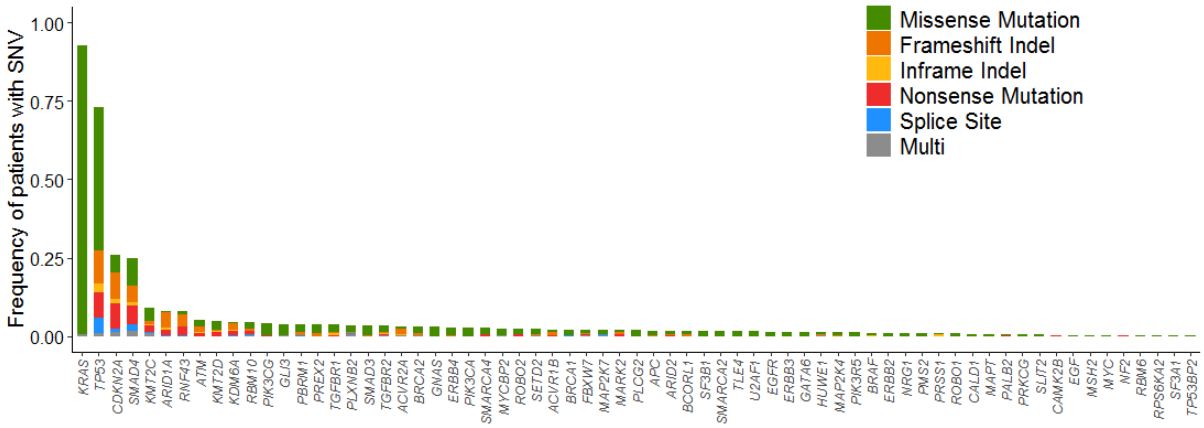


Figure 3.7: Bar plot of all SNVs in 293 PDAC patients. Frequency and type of all SNVs within the 67 genes investigated for mutations and small indel. Abbreviations: Multi = multiple mutation types within the same gene and patient. Previously published in Hoyer et al.[1]

At least one mutation was found in 98% of all patients (n=287), with a mean of 3.7 ± 2.0 SNVs per patient (Supplemental Figure 3). The four most commonly mutated genes were *KRAS* (93%), *TP53* (74%), *CDKN2A* (27%), and *SMAD4* (27%), with mutations targeting known hot spots (e.g., *KRAS* G12; Figure 3.8). Mutation types and mutation frequencies were comparable to previous PDAC sequencing studies (Supplemental Figure 4).

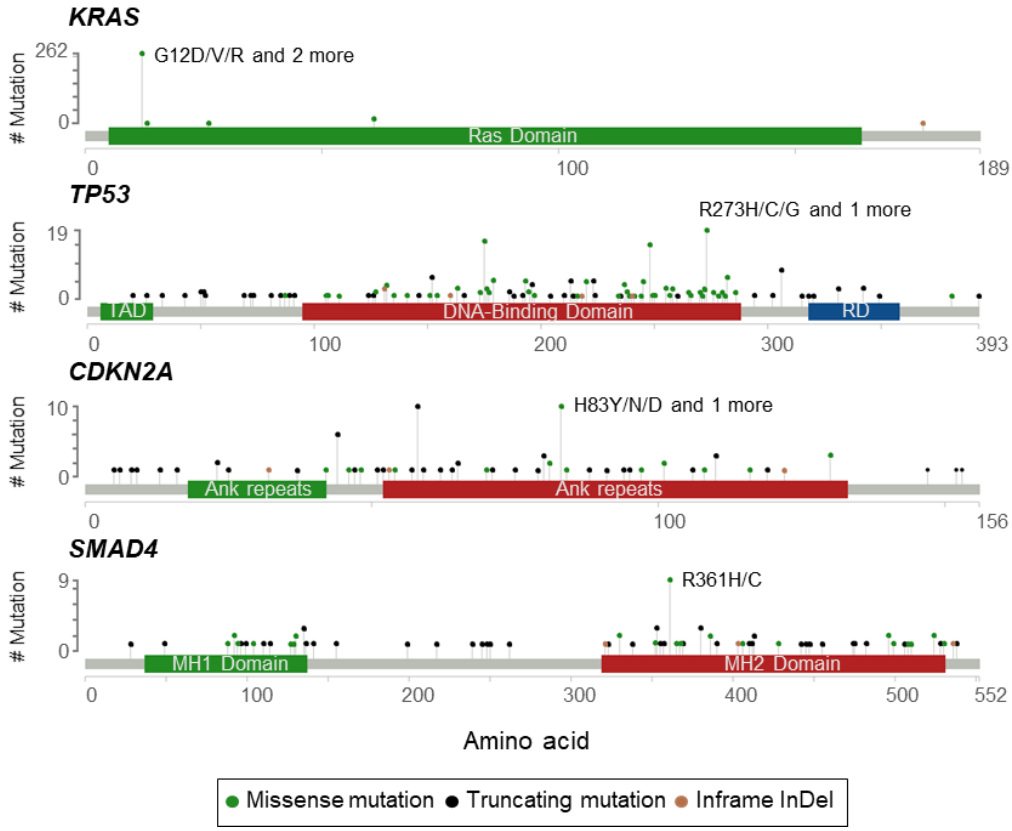


Figure 3.8: Lollipop plots of the four most commonly mutated genes. The most common SNVs are annotated for each gene. Abbreviations: TAD = Trans-activation domain, RD = Regulatory domain, Ank = Ankyrin, MH1/2 = MAD homology 1/2. Modified from cbioportal[167,168] and published in Hoyer et al.[1]

In accordance with the four most frequently mutated genes, the pathways most affected by SNVs were the MAPK pathway (93%), the cell cycle (76%), and the TFGβ signalling pathway (36%; Table 3.3).

Table 3.3: List of pathways affected by SNV in genes included in the PDAC panel. Each patient with a mutation in one of the genes and associated with the respective pathway is counted as affected in the pathway. Pathway annotations reported according to KEGG and Bailey *et al*.

Function	Patients		Genes included
	n	%	
MAPK Pathway	273	93.2 %	*BRAF, KRAS, MAP2K4, MAP2K7, MAPT, RPS6KA2*
Cell Cycle	223	76.1 %	*CDKN2A, MYC, TP53, TP53BP2*
TGFβ Signalling	107	36.5 %	*ACVR1B, ACVR2A, SMAD3, SMAD4, TGFBR1, TGFBR2*
Chromatin	53	18.1 %	*KDM6A, KMT2C, KMT2D, SETD2*
Others	48	16.4 %	*CALD1, GATA6, GLI3, GNAS, HUWE1, MARK2, PLCG2, PLXNB2, PRKCG, PRSS1*
SWI/SNF	46	15.7 %	*ARID1A, ARID2, PBRM1, SMARCA2, SMARCA4*
DNA Repair	36	12.3 %	*ATM, BRCA1, BRCA2, MSH2, PALB2, PMS2*
PI3K/Akt Pathway	33	11.3%	*PIK3CA, PIK3CG, PIK3R5, PREX2*
WNT Signalling	31	10.6%	*APC, CAMK2B, RNF43, TLE4*
RNA Processing	23	7.8%	*RBM10, RBM6, SF3A1, SF3B1, U2AF1*
EGFR Signalling	22	7.5%	*EGF, EGFR, ERBB2, ERBB3, ERBB4, NRG1*
ROBO SLIT	19	6.5%	*MYCBP2, ROBO1, ROBO2, SLIT2*
NOTCH Signalling	11	3.8%	*BCORL1, FBXW7, NF2,*

3.1.3.2 Copy number alterations in R0-resected PDAC

The strong fragmentation of FFPE DNA and the high stroma content of PDAC samples complicate CNA calling. For this reason, I used a second independent method to validate the CNAs found with the targeted NGS panel and to call CNA for samples too fragmented for NGS-based CNA calling. Of the 293 patients included in this analysis, 171 could be used for NGS-based CNA calling in the 100 genes included in the panel. The alteration frequencies detected in those patients were used to decide which genes to include in the MLPA analysis. Using the MLPA analysis, I validated 11 genes in the 171 good-quality samples, and I called CNAs in 112 patients that could not be analysed solely with NGS data.

A total of 4,157 CNAs were identified in 283 patients (Supplemental Table 5). In 80% of all patients (n=228), at least one genetic alteration was found, with a mean of 14.2 ± 14.1 CNAs per patient (Supplemental Figure 3). The number of alterations as well as their total copy number (tCN) burden were not associated with the morphological estimation of tumour content (Figure 3.9).

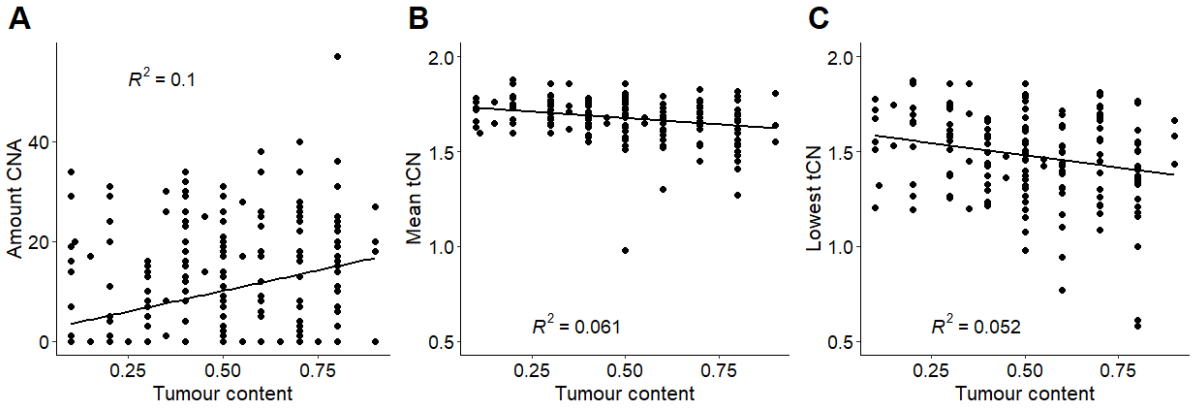


Figure 3.9: Comparison of tumour content with CNA characteristics. No correlation was found between morphological estimation of tumour content and **A)** amount CNA per patient, **B)** mean tCN of all variants in one patient, or **C)** lowest tCN**.** Both B and C include only deletions, therefore the signal is the strongest when it is lower (further away from normal tCN of 2). Modified from Hoyer et al.[1]

The major CNA loci were identified in genomic regions encoding for *SMAD4* (49%), *CDKN2A* (47%), and *TP53* (38%) as well as the 9p24 locus containing the immune checkpoint regulators *CD274*, *PDCD1LG2*, and *JAK2* (35%). Most CNAs were deletions (70%), though several well-known proto-oncogenes were found to be predominantly amplified (*KRAS*, *GATA6*, *MYC*) (Figure 3.10). These amplifications comprised about 24% of CNA, while the remaining 6% were UPD.

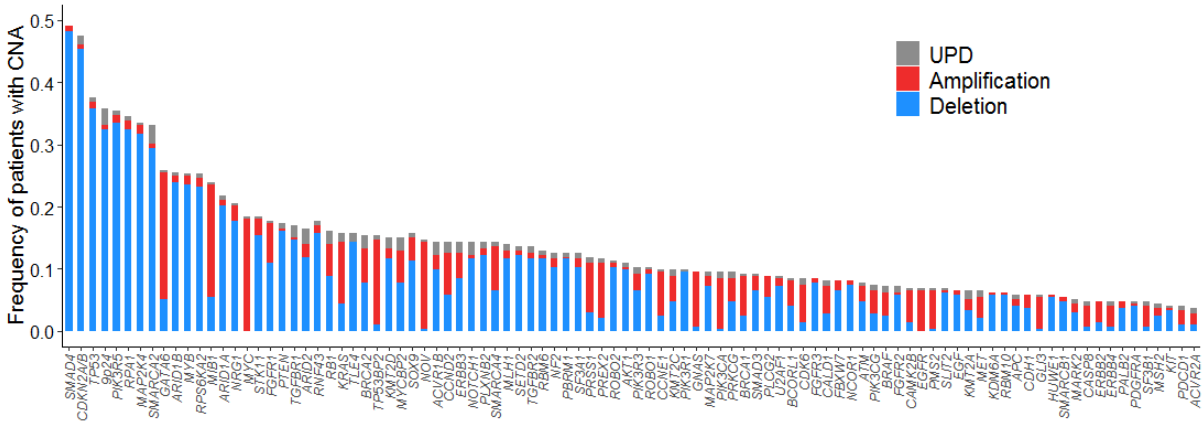


Figure 3.10: Bar plot of all CNAs in 283 PDAC patients. Frequency and type of all CNAs within the 100 genes investigated for copy number alterations. Previously published in Hoyer et al.[1]

A specific alteration pattern was observed in many patients and differed widely by chromosomal region. Gene-specific focal CNAs such as deletion of 9p21 with *CDKN2A/B* and gain of 18q11 with *GATA6* were identified, as were recurrent CNAs of entire chromosomes or chromosome arms like loss of chromosome 6 and gain of 1q (Figure 3.11).

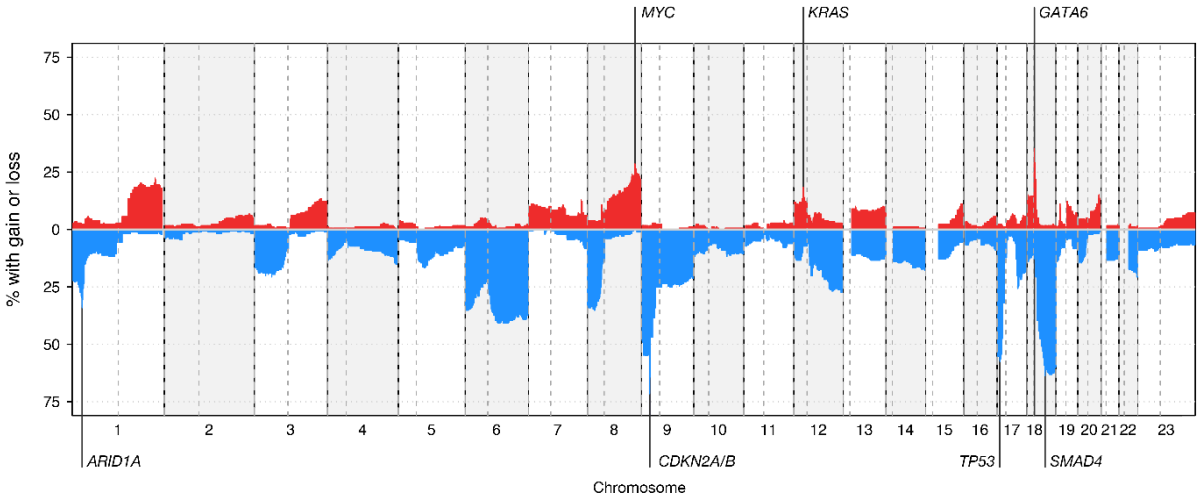


Figure 3.11: Genome-wide CNAs in 171 patients with complete copy number information. Example genes for most frequently altered regions are highlighted. Previously published in Hoyer et al.[1]

Different from the SNVs, which were predominantly found in two main pathways, CNAs were more evenly distributed across pathways, including cell cycle (74%), TGFβ signalling (64%), and PI3K/Akt signalling (63%; Table 3.4).

Table 3.4: List of pathways affected by CNAs with genes included in the PDAC panel. Each patient with a CNA in one of the genes, associated with the respective pathway was counted as affected by the pathway. Pathway annotations according to KEGG and Bailey et al.

Function	Patients		Genes included
	n	%	
Cell Cycle	210	74 %	*CCND2, CCNE1, CDK6, CDKN2A, CDKN2B, MYC, RB1, TP53, TP53BP2*
TGFβ Signalling	181	64 %	*ACVR1B, ACVR2A, SMAD3, SMAD4, TGFBR1, TGFBR2*
PI3K/Akt Pathway	178	63 %	*AKT1, MYB, PIK3CA, PIK3CG, PIK3R1, PIK3R3, PIK3R5, PREX2, PTEN, STK11*
SWI/SNF	155	55 %	*ARID1A, ARID1B, ARID2, PBRM1, SMARCA2, SMARCA4*
DNA Repair	155	55 %	*ATM, BRCA1, BRCA2, MLH1, MSH2, PALB2, PMS2, RPA1*
Others	151	53 %	*CALD1, CASP8, CDH1, GATA6, GLI3, GNAS, HUWE1, MARK2, PLCG2, PLXNB2, PRKCG, PRSS1, SOX9*
MAPK Pathway	150	53 %	*BRAF, FGFR1, FGFR2, FGFR3, KIT, KRAS, MAP2K4, MAP2K7, MAPT, MET, PDGFRA, RPS6KA2*
NOTCH Signalling	135	48 %	*BCORL1, FBXW7, MIB1, NF2, NOTCH1, NOV*
Chromatin	119	42 %	*KDM6A, KMT2A, KMT2C, KMT2D, NCOR1, SETD2, SMARCB1*
Immune Suppression	118	42 %	*CD274, JAK2, PDCD1, PDCD1LG2*
EGFR Signalling	113	40 %	*EGF, EGFR, ERBB2, ERBB3, ErbB4, NRG1*
WNT Signalling	101	36 %	*APC, CAMK2B, RNF43, TLE4*
RNA Processing	89	31 %	*RBM10, RBM6, SF3A1, SF3B1, U2AF1*
ROBO SLIT	74	26 %	*MYCBP2, ROBO1, ROBO2, SLIT2*

Major differences to altered pathways could be seen between deletions and amplifications. While players in the PI3K/Akt, the TGFβ signalling, and the SWI/SNF chromatin-remodelling pathways were almost always deleted, cell-cycle modifiers, EGFR-signalling molecules, and parts of the NOTCH signalling cascade were found to be amplified equally as often as they were deleted (Figure 3.12).

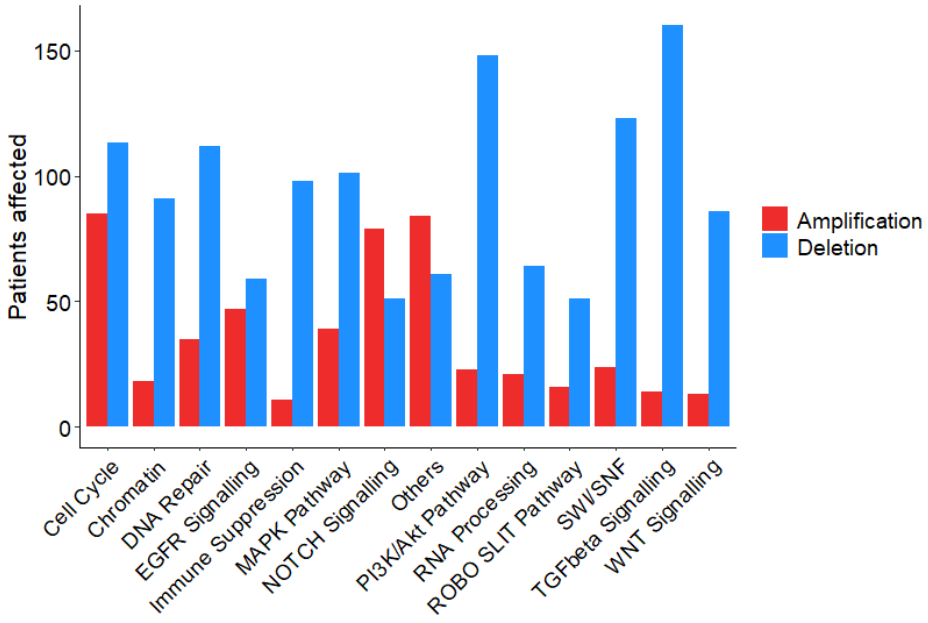


Figure 3.12: Number of patients with alteration in one of the 13 pathways grouped by alteration type. Some pathways are mostly affected by gene deletions, while others contain both amplification and deletions at a comparable rate.

3.1.4 Genetic landscape of PDAC

Combining SNVs and CNAs, I achieved a comprehensive picture of the genetic background of R0-resected PDAC patients. In 99% of all patients (n=290), at least one genetic alteration was found, with the PDAC genes (*KRAS,TP53, CDKN2A* and *SMAD4*) being altered by both mutations and small indels as well as larger deletions or amplifications (Figure 3.13).

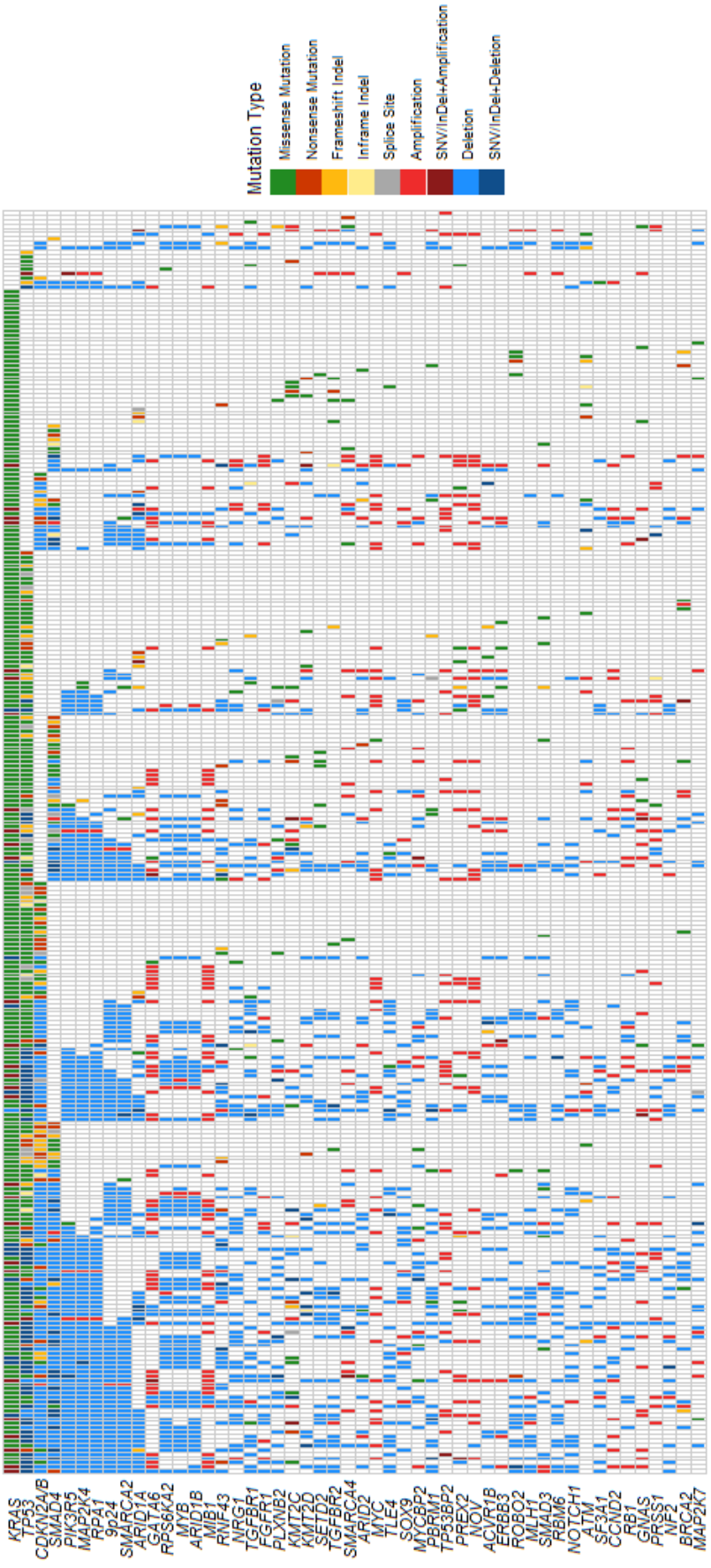


Figure 3.13: Overview of genetic alterations in R0-resected PDAC patients. Landscape plot of the 50 most frequently altered genes in 293 PDAC patients. Modified from Hoyer et al.[1]

The main pathways affected by SNVs and CNAs were the MAPK-pathway (99%), cell cycle control (92%) and TGFβ signalling (77%). Genes encoding for members of the PI3K/Akt pathway (65%) and genes involved in chromatin remodelling (71%) were also frequently affected in the CO-2016 cohort (Figure 3.14,

Supplemental Table 6). The *EGFR* gene was found to be mutated in four patients (1.4%) and amplified in 19 patients (6.5%).

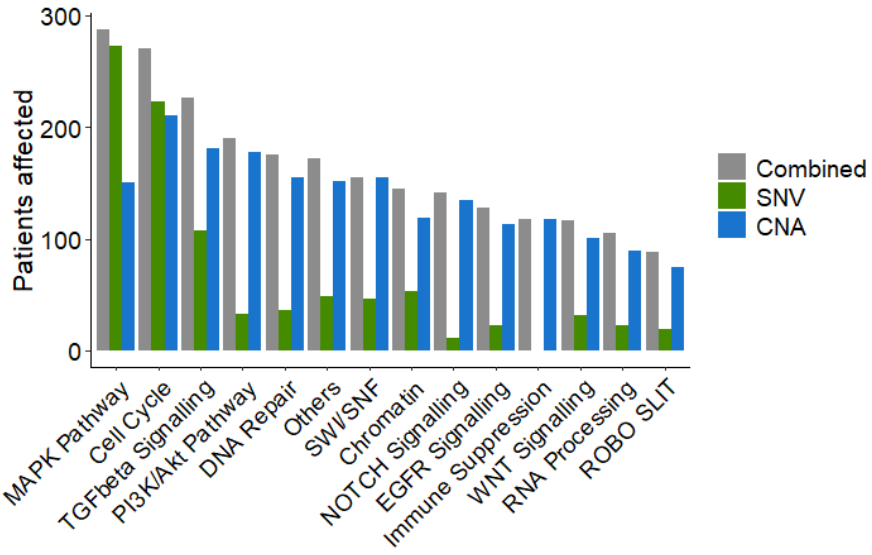


Figure 3.14: Overview of all pathways altered by either SNV or CNA or a combination of both. The MAPK pathway is mostly altered by SNVs, while cell cycle signalling is affected equally as often by SNV and CNA. All the remaining functional groups, including TGFβ signalling, have more CNAs, than SNVs.

To identify patterns within in the CO-2016 cohort, which might reveal biologically distinct subgroups, I searched for pairwise gene associations using fisher's exact test. A total of 14 pairs were significantly associated with a false discovery rate <5 %. The three major CNAs affecting the genes *SMAD4*, *CDKN2A*, and *TP53* showed strong co-occurrences, which means that many patients had two or more of these genes concomitantly deleted. In addition, *TP53* mutations were often accompanied by mutations in *KRAS*-, and *CDKN2A/B* as well as CNAs in *SMAD4*. While mutually exclusive gene pairs often imply functional redundancy, such a pattern was rarely observed. Only *ATM* mutations showed little co-occurrence with *TP53* mutations, especially with *TP53* activating mutations (Figure 3.15).

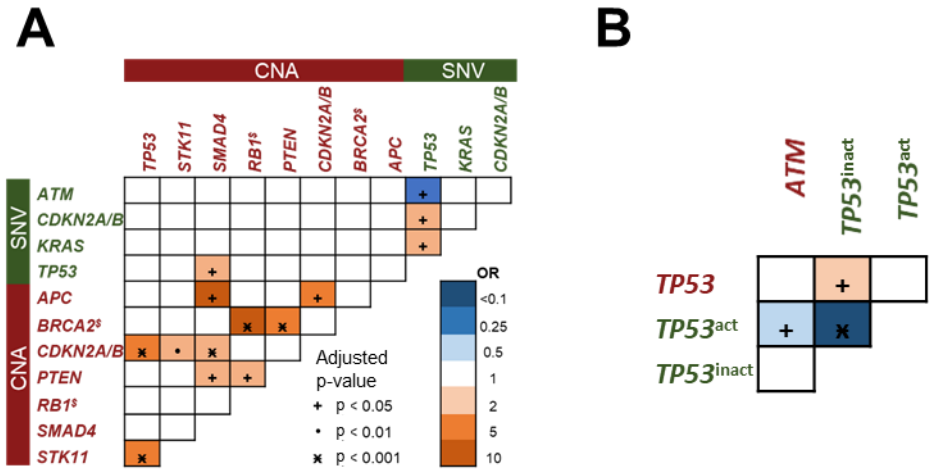


Figure 3.15: Co-occurrence and mutual exclusivity plot for all 293 patients. **A)** Genes mutated in at least 4 % of patients and CNAs in all MLPA genes were included in the analysis. Significance levels of multiple testing adjusted p-values are shown with symbols, OR with colours (blue shows different levels of mutual exclusivity, oranges show different levels of co-mutation). **B)** Co-occurrence and mutual exclusivity test for TP53 mutations. For a list of all *TP53* mutations with functional consequences see Supplemental Table 7. Co-occurrence or mutual exclusivity was determined through fisher's exact test. All p-values were multiple testing corrected via the Benjamini–Hochberg method. Abbreviations: act = activating mutations, inact = inactivating mutations, $ = genes are located within same chromosomal region on chromosome 13q14. Modified from Hoyer et al.[1]

.

3.2 Impact of single gene alterations on clinical course

To find genetic alterations that might be used as prognostic biomarkers for patients' risk stratifications or to identify a subgroup that benefits from erlotinib treatment, I analysed SNV, can, and expression profiles for their individual correlations with clinical outcome and erlotinib treatment success.

3.2.1 Clinical background criteria in the CO-2016 cohort

The 293 patients in the CO-2016 cohort had a median age of 64 years, with the youngest patient being 24 and the oldest 82 years old. The cohort had a slight male surplus (56%) and was well balanced in terms of treatment arm distribution (51% Gem vs 49% Gem+Erlo). The vast majority of patients had good clinical performance, with two-thirds of all patients having a Karnofsky score of 90 or higher. The median CA 19-9 level was 19.5, ranging from 1 to 5,816. Fewer than 20% of patients had elevated CA 19-9 values (<100). The resected tumours were mostly moderately (61%) or poorly (31%) differentiated and had a T-staging of 3 (86%). In two thirds of all patients, the tumour had spread to 1–3 lymph nodes in close proximity to the pancreas (N1), while the remaining patients had no lymph node involvement (Table 3.5).

Table 3.5: Clinical characteristics of the CO-2016 cohort. Previously published in Hoyer et al.[1]

Characteristics	PDAC
	(n=293)
Age, years	
Median	64
Range	24-82
Sex	
male - no. (%)	163 (56 %)
female - no. (%)	130 (44 %)
Arm	
Gemcitabine - no. (%)	149 (51 %)
Gemcitabine+Erlotnib - no. (%)	144 (49 %)
Karnofsky	
60 - no. (%)	1 (>1 %)
70 - no. (%)	10 (3 %)
80 - no. (%)	75 (26 %)
90 - no. (%)	112 (38 %)
100 - no. (%)	95 (32 %)
Grading	
G1 - no. (%)	7 (2 %)
G2 - no. (%)	180 (61 %)
G3 - no. (%)	96 (33 %)
unknown - no. (%)	10 (4 %)
T-Stage	
T1 - no. (%)	9 (3 %)
T2 - no. (%)	30 (10 %)
T3 - no. (%)	251 (86 %)
T4 - no. (%)	3 (1 %)
N-Stage	
N0 - no. (%)	106 (36 %)
N1 - no. (%)	187 (64 %)
Postoperative CA 19-9, kU/L	
Median (range)	19.5 (1-5816)
≤ 100 - no. (%)	223 (76 %)
101-500 - no. (%)	29 (10 %)
> 500 - no. (%)	12 (4 %)
unknown - no. (%)	29 (10 %)

All baseline characteristics, OS and DFS, and the impact of erlotinib treatment were similar to the entire CONKO-005 cohort (Figure 3.16 A). As in the CONKO-005 trial, no significant increase in OS or DFS was observed when adding erlotinib to gemcitabine as adjuvant therapy[142]. A late split in the curves might indicate a small sensitive population (Figure 3.16 B). Altogether, the CO-2016 cohort appeared a good representation of the CONKO-005 trial cohort.

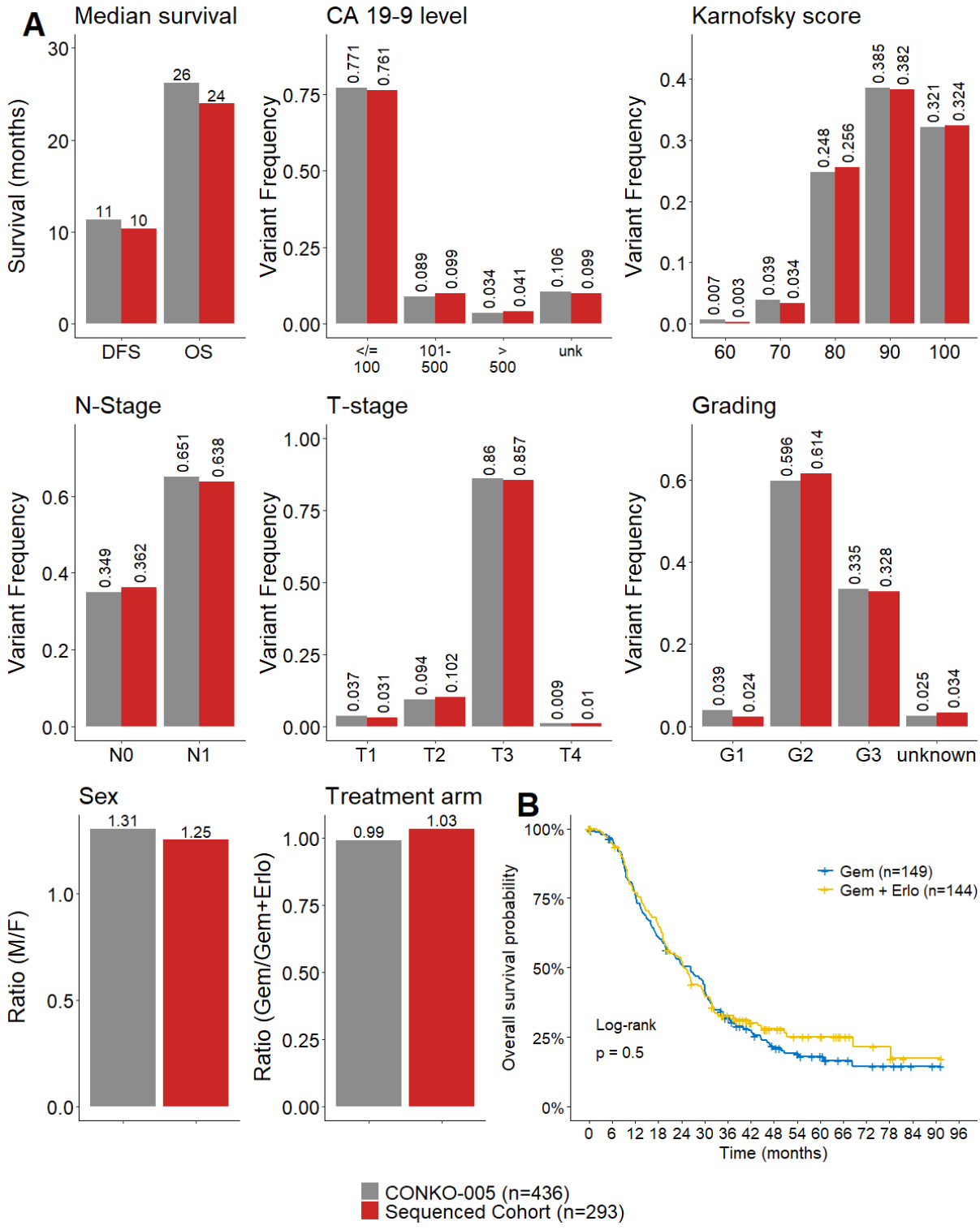


Figure 3.16: Overview of clinical baseline characteristics and course of disease. A) Comparison of baseline characteristics between the entire CONKO-005 study population (grey) and the subset of 293 patients sequenced within this study (red), indicating a high comparability of both patient groups. **B)** The Kaplan-Meier estimate of OS of the 293 sequenced PDAC patients according to the two treatment arms of the CONKO-005 trial: Gemcitabine (blue) vs. Gemcitabine + Erlotinib (yellow). Previously published in Hoyer et al.[1]

In the treatment of PDAC patients, several clinical characteristics (such as N-Stage, grading, and Karnofsky score) are being used for outcome prediction and subsequent treatment

decisions. I analysed which of these clinical characteristics influenced DFS and OS in the CO-2016 cohort. In a first univariate cox regression analysis, three clinical variables showed significant impact on DFS and OS: N-staging, postoperative CA 19-9 levels, and grading. This indicates that all three clinical variables have a significant impact on OS and DFS, to an extent specified by their hazard ratios (HR; Figure 3.17).

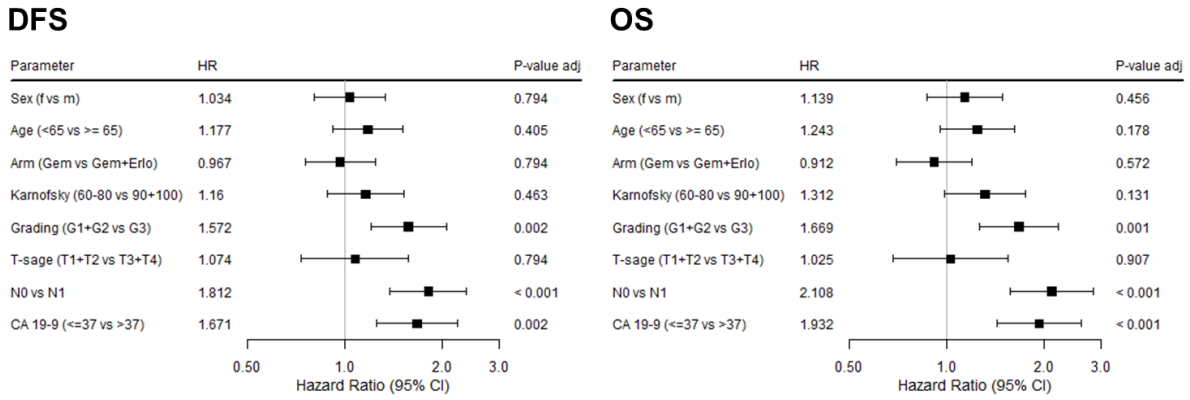


Figure 3.17: Forest plot of univariate cox regression analysis including all clinical baseline characteristics. Parameters were dichotomised as indicated. For CA 19-9 levels, the clinical cut-off of 37 U/ml was used to sort patients into either normal or increased CA 19-9 levels, as is done in clinical practice[169]. All p-values were adjusted for multiple testing. The squares show HR and the arms 95% confidence intervals (CI).

3.2.2 Single gene analysis on survival and relapse pattern

3.2.2.1 Impact of SNVs on course of disease and erlotinib sensitivity

To find a single genetic alteration or a combination of several genetic alterations with predictive or prognostic power, I first tested each alteration independently using a univariate cox regression analysis. For SNVs, I tested the 67 panel genes. Three of the most commonly mutated genes (*KRAS*, *TP53*, and *SMAD4*) were further subdivided according to specific amino acid changes (*KRAS* G12, G12V, G12D, G12R, Q61) or their functional consequences (non-sense – stop gain and frameshift indel – and missense). None of the 76 variants tested significantly impacted DFS or OS after adjusting for multiple testing (Figure 3.18).

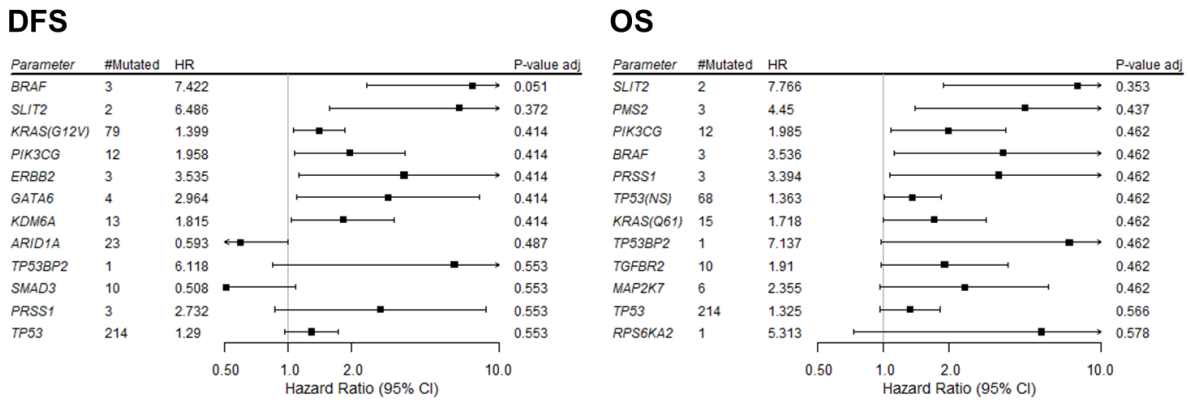


Figure 3.18: Forest plot of univariate cox regression analysis for all 67 panel genes. *KRAS* was further sub-grouped into amino acid changes (G12, G12V, G12D, G12R, Q61) and *SMAD4,* and *TP53* into was sub-grouped into functional SNV groups (NS = non-sense [stopgain and frameshift indel] and MS = missense). Only genes with an unadjusted p-value of <0.1 are shown. #Mutated = number of patients with a gene mutation; P-value adj = p-value adjusted for multiple testing using the Benjamini and Hochberg method[145].

Next, I used the same SNV categories to calculate interaction p-values comprising the mutation status and the treatment arm. The test for interaction shows whether patients with a specific SNV would benefit from additive erlotinib treatment. Only one SNV had significant impact on erlotinib sensitivity after multiple testing corrections: the Phospholipase C Gamma 2 (PLCG2), a transmembrane signalling enzyme that catalyses the conversion of 1-phosphatidyl-1D-myo-inositol 4,5-bisphosphate to 1D-myo-inositol 1,4,5-trisphosphate (IP3) and diacylglycerol (DAG) using calcium as a cofactor. Patients with a mutation in this gene had a significantly decreased HR for both OS and DFS when treated with erlotinib (Figure 3.19). It should be noted that the tests' impact is weakened by the exceptionally low mutation frequency of *PLCG2*, which was mutated in only six patients and only in one patient with gemcitabine treatment (Supplemental Figure 5).

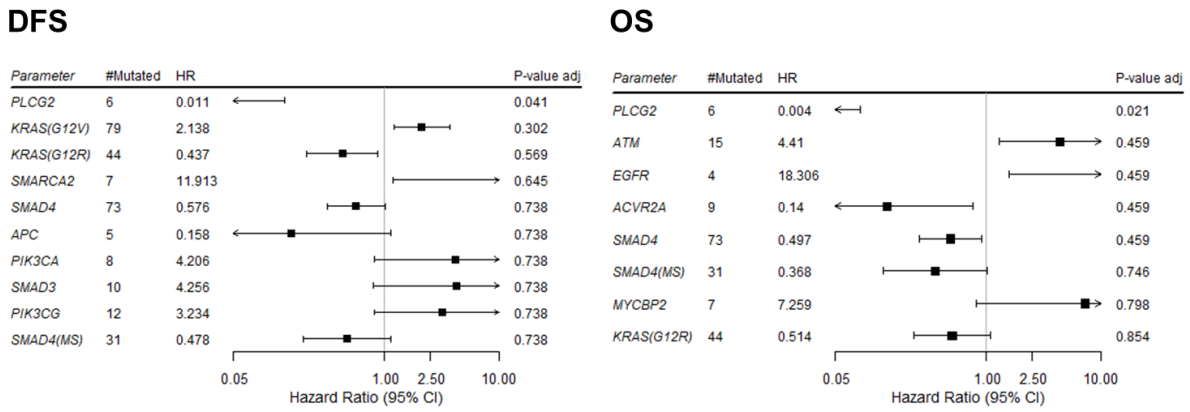


Figure 3.19: Forest plot of univariate cox regression analysis testing interaction of mutation status and treatment arm. *KRAS* was further sub grouped into amino acid changes (G12, G12V, G12D, G12R, Q61) and *SMAD4* and *TP53* into functional SNV groups (NS = non-sense [stop gain and frameshift indel] and MS = missense). Only genes with an unadjusted p-value of <0.1 are shown. #Mutated = number of patients with a mutation in the gene; P-value adj = interaction p-value adjusted for multiple testing using the Benjamini and Hochberg method[145].

Enrichment analysis comparing different relapse pattern (local vs distance metastasis pattern) did not reveal any significant enrichment of SNVs in either subgroup.

3.2.2.2 Impact of CNAs on course of disease and erlotinib sensitivity

To detect CNAs with potential impact on OS, DFS, or erlotinib sensitivity, I analysed all 11 genes covered by the MLPA panel in 283 patients using the same method as that for SNVs. Similarly, no alteration was significantly associated with increased or decreased DFS or OS after multiple testing corrections (Figure 3.20).

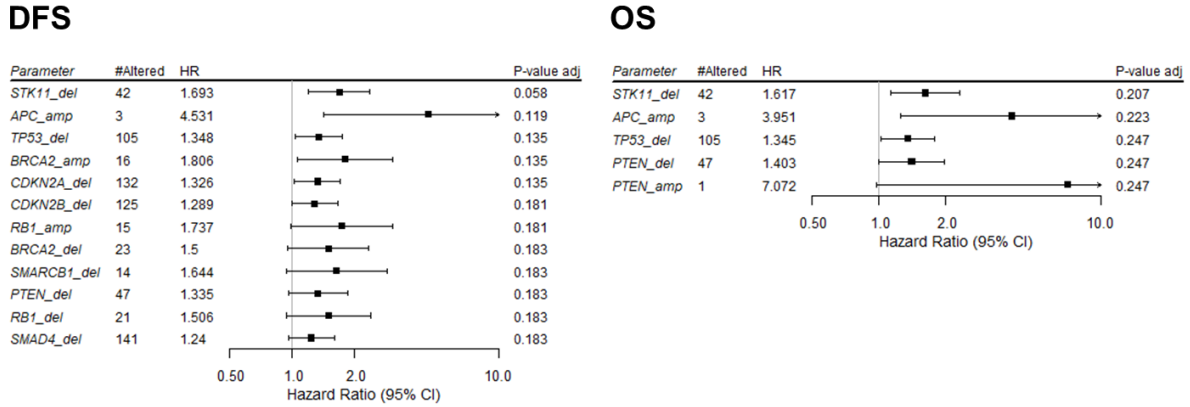


Figure 3.20: Forest plot of univariate cox regression analysis for all 11 MLPA validated genes. Only genes with an unadjusted p-value <0.1 are shown. #Altered = number of patients with an alteration in the gene; P-value adj = p-value adjusted for multiple testing using the Benjamini and Hochberg method[145].

Only one genetic alteration showed significant impact on erlotinib sensitivity after correcting for multiple testing. Patients with amplification in tumour suppressor *STK11* had significantly decreased HR when treated with additive erlotinib as compared to the gemcitabine treatment

arm (Figure 3.21). As with *PLCG2* mutations, this finding should be evaluated carefully due to its low deletion frequency and the fact that only one *STK11*-amplified patient could be found in the gemcitabine treatment arm (Supplemental Figure 6).

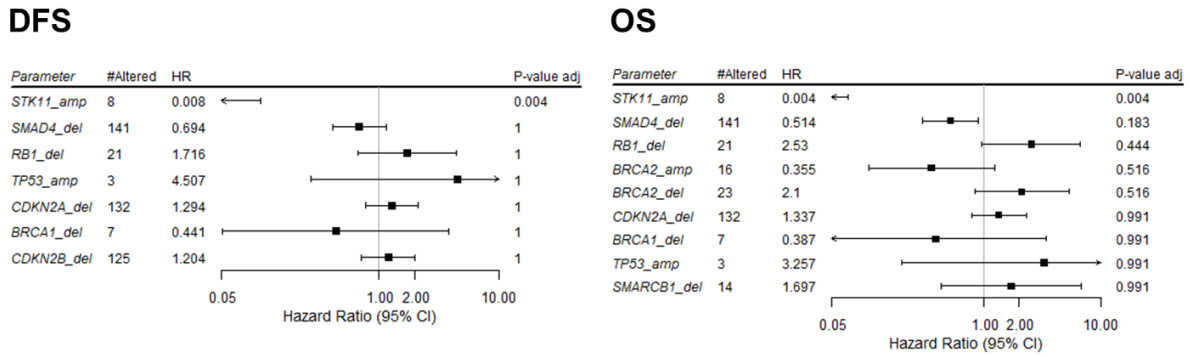


Figure 3.21: Forest plot of univariate cox regression analysis testing interaction of CNA status and treatment arm. Only genes with an unadjusted p-value of <0.5 are shown. #Altered = number of patients with a mutation in the gene; P-value adj = interaction p-value adjusted for multiple testing using the Benjamini and Hochberg method[145].

Enrichment analysis comparing different relapse patterns (local vs distance metastasis patterns) did not reveal any significant enrichment of CNAs in either subgroup.

3.2.2.3 Impact of gene expression on course of disease and erlotinib sensitivity

Applying the same approach to the 770 genes included in the NanoString Panel resulted in the mRNA expression levels of 19 genes to be significantly associated with OS and 17 with DFS. In all cases, high expression levels after dichotomisation in high and low expression led to increased HR. To decrease the number of variables for further subsequent multivariate analyses, a preselection round was performed by placing three significant clinical variables (Karnofsky score, grading, N-stage) into individual multivariate cox regression analysis for each gene. This led to the discovery of six genes with significant associations with OS and five with DFS (Figure 3.22).

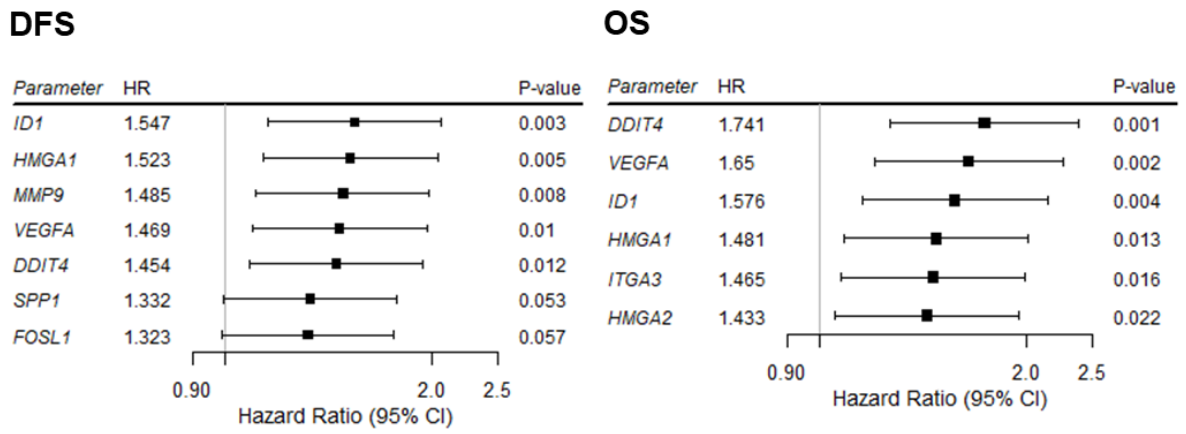


Figure 3.22: Individual multivariate cox regression analysis for single gene expression level. For each gene, a single multivariate cox regression containing three clinical variables and the gene of interest was performed. Genes were dichotomized along their medians (low vs high expression). In all cases, high expression values were associated with higher HR.

As a second independent test, I used the MaxStat analysis to identify genes containing expression levels with a significant impact on DFS or OS. The MaxStat analysis can be used to adjust the cut-off value between high and low expression in terms of the highest possible split of the two curves in a Kaplan-Meier analysis. This ensures that smaller patient groups with very high or low expression values are not lost due to a strict split along the median expression. Subsequent adjustment for multiple testing led to the identification of four genes with significant impact on OS but none for DFS when highly expressed in patients (Figure 3.23).

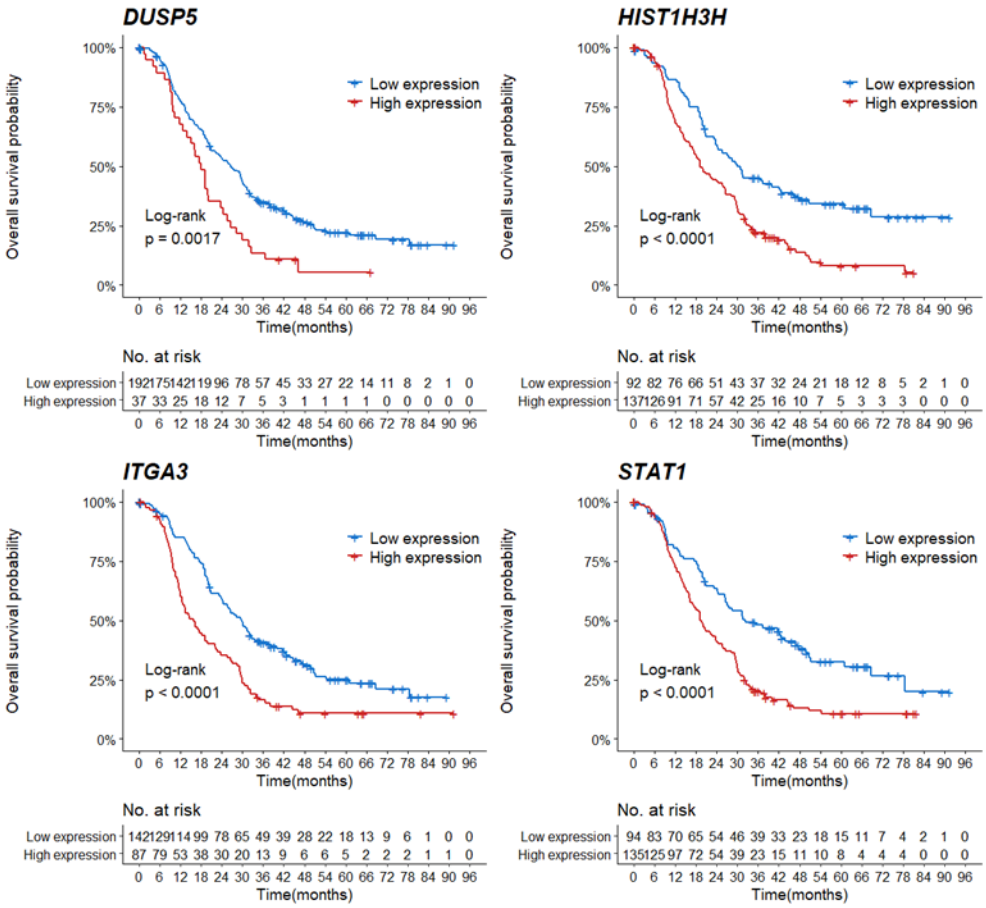


Figure 3.23: Kaplan-Meier curves of results from MaxStat analysis. Shows all genes with expression levels significantly associated with OS after multiple testing in the MaxStat analysis. The cut-off between high and low expression was adjusted for each gene individually to achieve maximal split in the curve. Expression cut off values were DUSP5 = 487 counts, HIST1H3H = 620 counts, ITGA3 = 1069 counts, and STAT1 = 517 counts.

3.2.3 Genetic risk score for survival prediction in PDAC

Currently, no established risk models are available for PDAC, and individual risks are mainly estimated on individual physicians' experiences. As described above, I identified several clinical- and expression-based variables with an independent impact on DFS and OS. To combine these findings and thereby increase their predictability, I created a genetic risk score for resectable PDAC. I decided to base the genetic risk score on the expression values of the 230 patients with known expression statuses. Therefore, I repeated the univariate cox regression analysis of all clinical baseline characteristics for these 230 patients and found four clinical variables to have a significant impact on OS (grading, N-Stage, Karnofsky performance status, and CA19-9 level; Figure 3.24 A). Additionally, I choose to enter the five genes with high expression levels significantly associated with low OS into the individual multivariate analysis (*DDIT4, HMGA1, ID1, ITGA3, VEGFA*) as well as the four final results of the MaxStat analysis (*DUSP5, HIST1H3H, ITGA3, STAT1*; Figure 3.24 B).

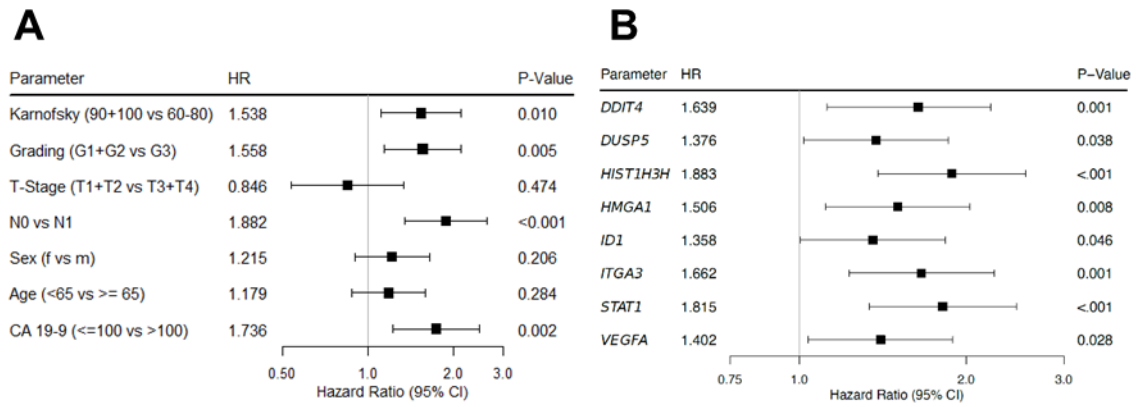


Figure 3.24: Univariate cox regression analysis of input values for risk score generation. A) Represents all clinical baseline characteristics tested within the 230 patients with available expression values. **B)** The five genes that were found to associate significantly with OS via single multivariate analysis plus the four genes from the MaxStat analysis (ITGA3 was significant in both). All gene expression levels were dichotomized using their median expression levels.

When combined in a multivariate cox-regression using backwards selection, two clinical variables (N-stage and grading) as well as mRNA expression levels of three genes (*DDIT4, HIST1H3H, ITGA3*) retained independent significance with a p-value of <0.05 (Figure 3.25). In line with our results, high expression of *DDIT4* – encoding a DNA-damage response regulator – as well as the cellular senescence regulating histone H3.1 (*HIST1H3H*), and the cell invasion promoting cell surface adhesion molecule Integrin alpha-3-expressing *ITGA3* all showed to be unfavourable prognostic markers in PDAC in the human protein atlas[170]. All five variables had an independent negative impact on OS with comparable HRs between 1.5 and 1.9 (Figure 3.25 A). Thus, no further weighting of variables was included in subsequent risk score development.

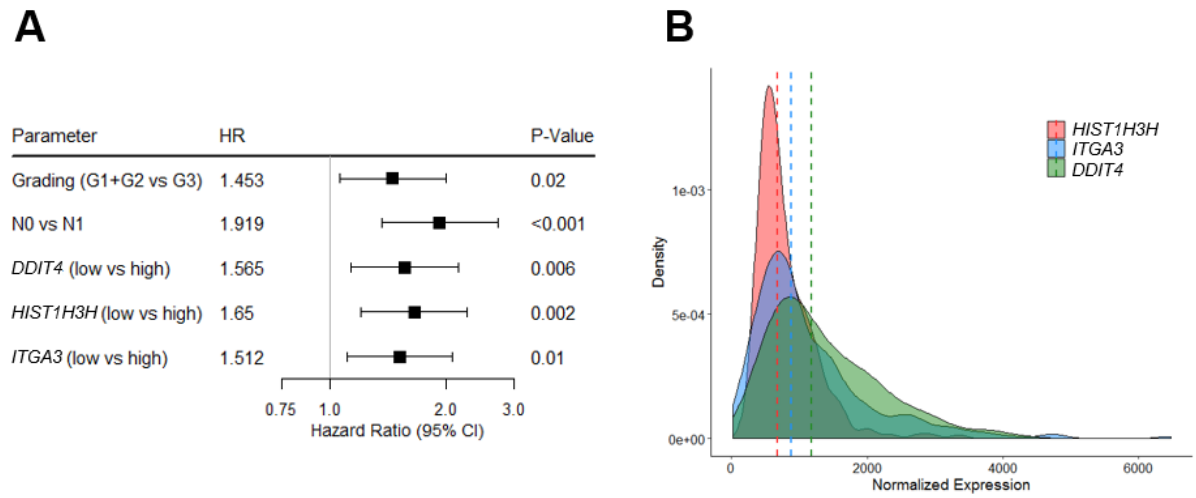


Figure 3.25: Overview of the genetic risk score. A) Multivariate cox regression analysis of the final five components of the risk score. **B)** Expression levels of the three genes (*HIST1H3H, ITGA3, DDIT4*) included in the molecular risk score. Dashed lines mark the median expression count which was used for high versus low expression dichotomisation.

I created a combined risk score by assigning each patient one point for each of the five variables with a negative effect on OS (N1, G3, overexpression of one or more of the three genes). Patients with 0–1 points were classified as 'low-risk' (n=63), 2–3 points as 'intermediate-risk' (n=115), and 4–5 points as 'high-risk' (n=51). The three risk groups showed significantly different OS and DFS (log-rank p<0.001) with median survival times of 50, 25, and 13 months (DFS: 20, 10 months, and 7 months) respectively (Figure 3.26). Thus, patients' survival times were halved from one risk group to the other, reaching survival times comparable to metastatic PDAC in high-risk R0-resected patients[90,91].

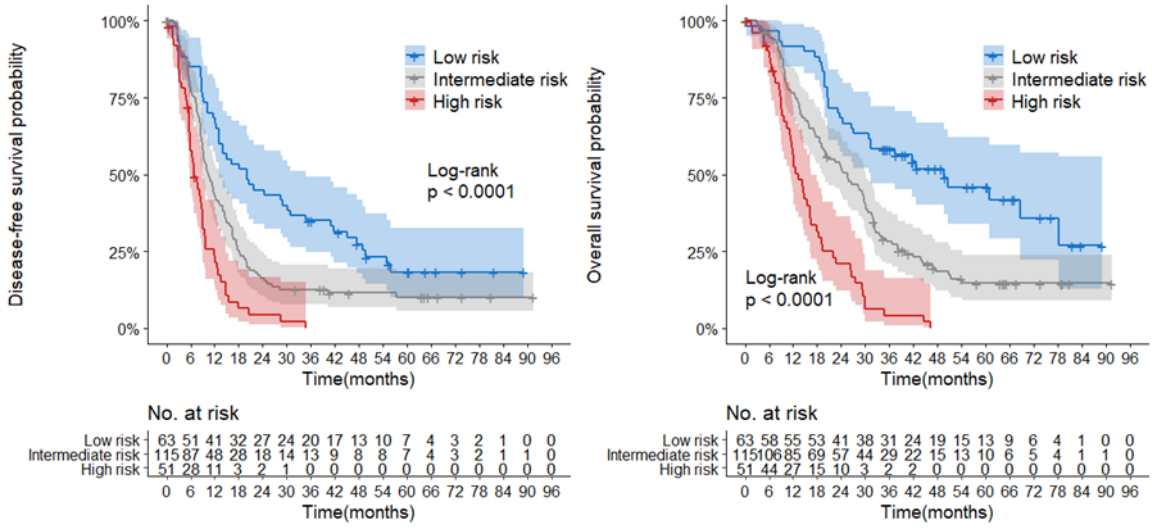


Figure 3.26: Kaplan-Meier curves of survival groups defined using the genetic risk score. DFS and OS of all three risk groups are show with their respective 95 % confidence intervals.

This coincided with subsequent metastasis patterns. Disease relapse occurred significantly less often in low-risk patients (47/63=75%) than in patients from high-risk groups (48/51=94%; two-sided Fisher's exact p=0.011; Figure 3.27 A). Additionally, the proportion of isolated local relapses was higher in low-risk group patients at 24% (15 local out of 47 relapses) compared to 12 % and 10 % in intermediate- and high-risk group patients (two-sided Fisher's exact p=0.006 and p=0.018, respectively; Figure 3.27 B). Finally, the low-risk group had significantly greater lung metastasis at 17% (metastases found only in the lung; 4 out of 23 distant metastases) as compared to 3% in the high-risk group (Fisher's exact p=0.041). This represents a clinical pattern associated with increased OS in PDAC[171].

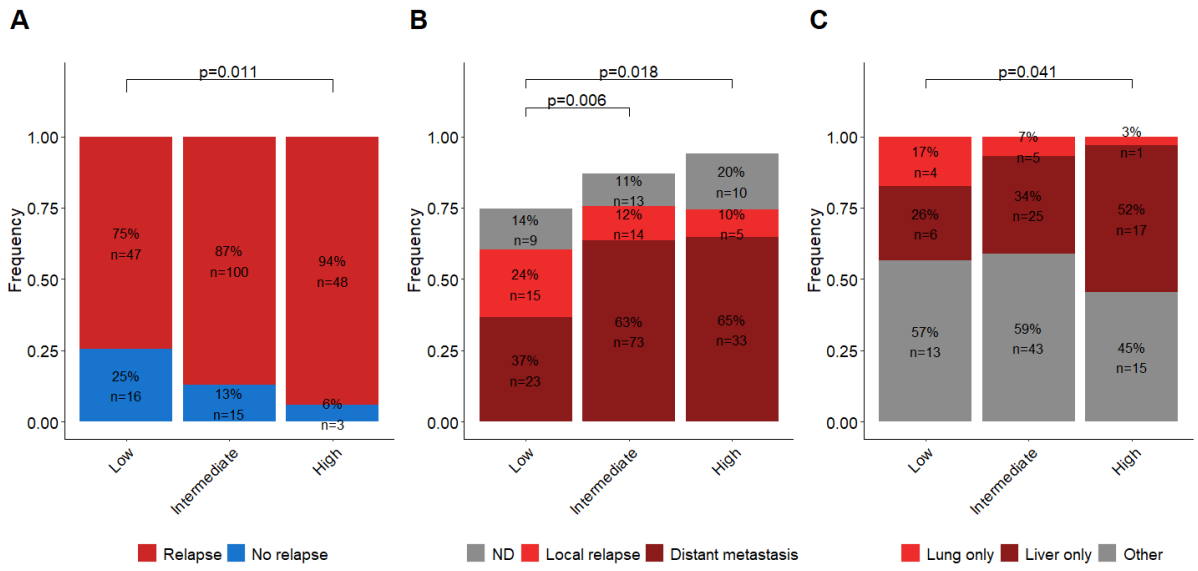


Figure 3.27: Metastasis pattern in survival score risk groups. A) Proportions of patients with or without relapse in each risk group. Total patient numbers as well as percentages from total patient numbers per risk group are depicted. **B)** Further subdivision of the relapse cases into either local or distant metastasis. Depicted are total patient number and percentages from total patient number per risk group. **C)** Shares of the different types of distant metastasis, dividing them into lung only (metastasis found only in the lung), liver only (metastasis found only in the liver), and other (distant metastasis pattern e.g., lung and liver, peritoneal metastasis). Total patient number and percentages from number of patients with distant metastases per risk group are also depicted. ND = no data.

To validate the risk score, I choose the patient cohort from Bailey et al.[102] as part of the ICGC[172] as a second independent validation group. Clinical data (AJCC tumour stage, grading, survival status, and OS) as well as normalized mRNA expression (*ITGA3*, *DDIT4*, *HIST1H3H*) were extracted from 96 patients with available data. Patients with AJCC stage III and IV as well as patients with tumour grading of 4 or no known tumour grade were excluded to ensure comparability with the CO-2016 cohort. A total of 85 patients matched these criteria and were subsequently classified into low-, medium-, and high-risk groups. The median of normalized expression counts was used for dichotomisation into high and low expression. As for the CONKO-005 trial cohort, half of the ICGC patients were assigned to the intermediate-risk group. Again, the three risk groups showed significantly different OSes (log rank p=0.004) with median survival rates of 12, 20, and 50 months (Figure 3.28 A). Similar to what could have been observed in the CO-2016 cohort, patients in the validation cohort showed significant differences in relapse pattern across the three risk groups. Disease relapse occurred significantly less often in patients in the low-risk groups (9/24=38%) as compared to patients in the intermediate-risk (25/41=61%; two-sided Fisher's exact p=0.03) and high-risk (14/20=70%; two-sided Fisher's exact p=0.01) groups (Figure 3.28 B).

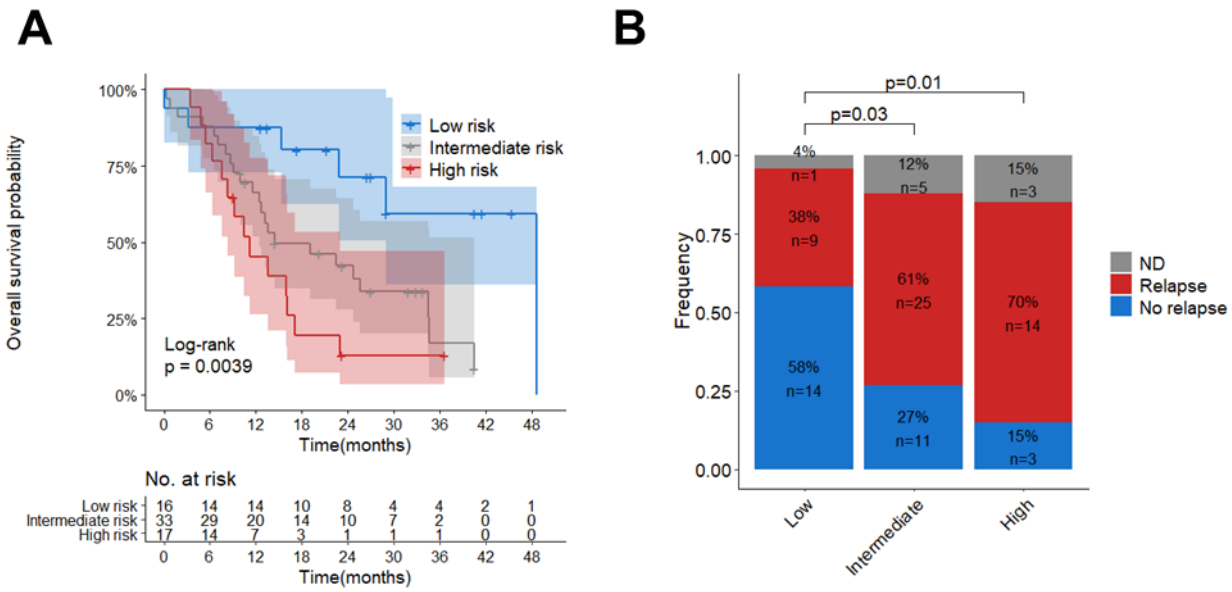


Figure 3.28: Overview of validation cohort after grouping with the genetic risk score. A) Kaplan-Meier analysis showing the significant survival differences among patients in the low- (blue), intermediate- (grey), and high-risk groups (red). **B)** Metastasis pattern in survival score risk groups. Proportions of patients with or without relapse in each risk group. Total patient numbers as well as percentages from total patient number per risk group are depicted.

3.3 Non-negative matrix factorization reveals molecular subgroups with distinct clinical outcomes

Because single genetic alterations failed to exhibit significant predictive or prognostic power, a more complex subgrouping approach became necessary. The NMF is a machine learning approach which can be used to find signatures in large data sets with sparse matrices as input. It has proven to be a powerful tool for pattern recognition in cancer mutations[173,174]. For my data set, I chose to input mutation status of all 67 panel genes and the alteration status of the 11 MLPA genes available for 283 patients. Hierarchical clustering on the signatures extracted with NMF led to five robust clusters (Figure 3.29).

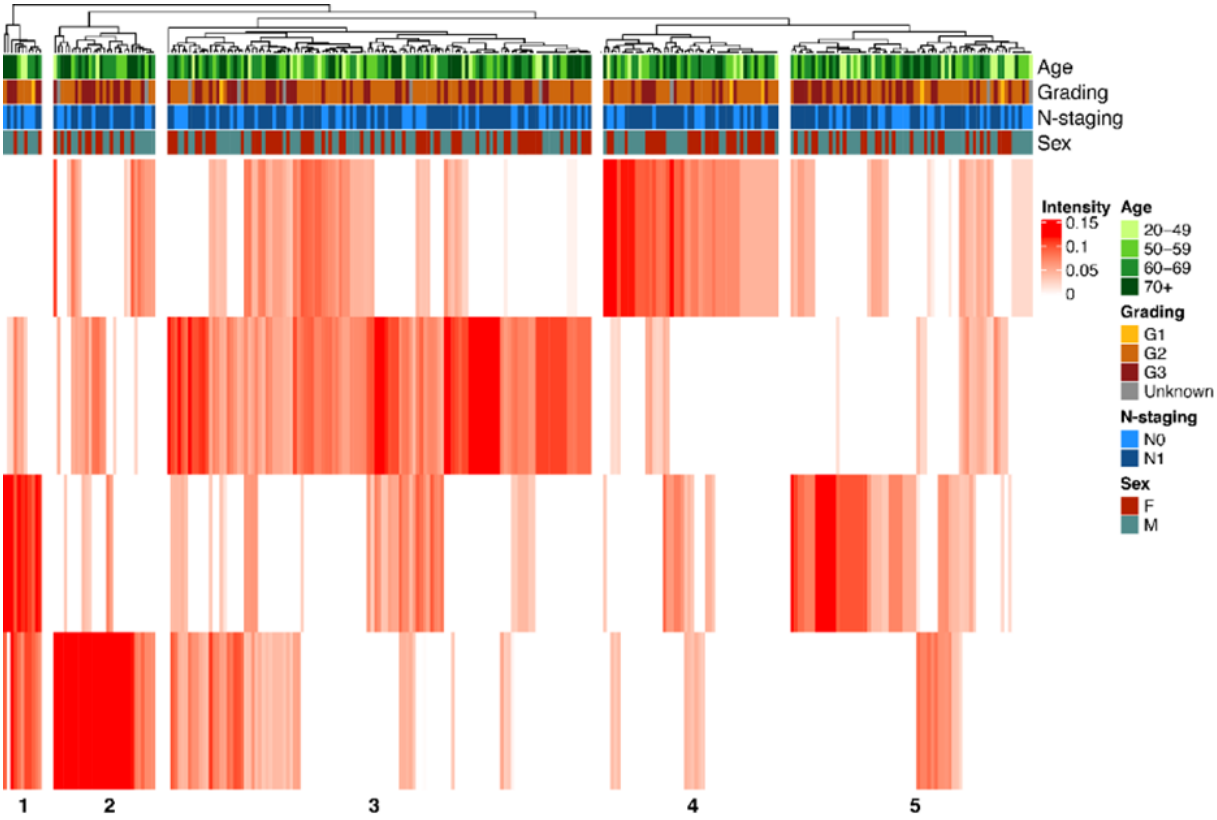


Figure 3.29: Unsupervised patient clustering using non-negative matrix factorization. Heat map of 283 patients, clustered based on four meta-genes with their respective clinical baseline characteristics. The patient subgroups are indicated below the heat map. Modified from Hoyer et al.[1]

3.3.1 Clinical and genetic characteristics of NMF cluster

While baseline characteristics such as age, gender, N-stage, and grading were distributed similarly across all clusters (Figure 3.29,

Supplemental Table 8), a wide range of distinct genetic and clinical features were attributed to each cluster. Tumours from patients in cluster 1 (n=11) had significantly more mutations (8.6 vs 3.5 in all other groups, two-sided Fisher's exact p=0.003) as well as a significant enrichment of mutations and amplifications in genes from the ERBB signalling pathway (*PLCG2*, *MAP2K7*, *ERBB4*, and *CAMK2B*). Cluster 2 (n=29) was enriched for copy-number deletions in well-known tumour suppressor genes (*RB1*, *BRCA2*, and *PTEN*), while cluster 3 (n=121) contained numerous deletions affecting major PDAC gene loci (e.g., *CDKN2A*/*B*, *TP53*, and 9p24). In contrast, clusters 4 (n=50) and 5 (n=69) both possessed much fewer alterations per tumour (12.2 and 9.2 vs 25.94, 24.4, and 22.4 in clusters 1, 2, and 3). In cluster 4, this was based mainly on the absence of *CDKN2A*/*B* deletions, whereas cluster 5 contained frequently altered genes including *SMAD4*, *CDKN2A*/*B*, *TP53*, and *KRAS* – all of which were affected less often. Additionally, a significant enrichment of *SMAD4* mutations in cluster 4 was observed (Figure 3.30).

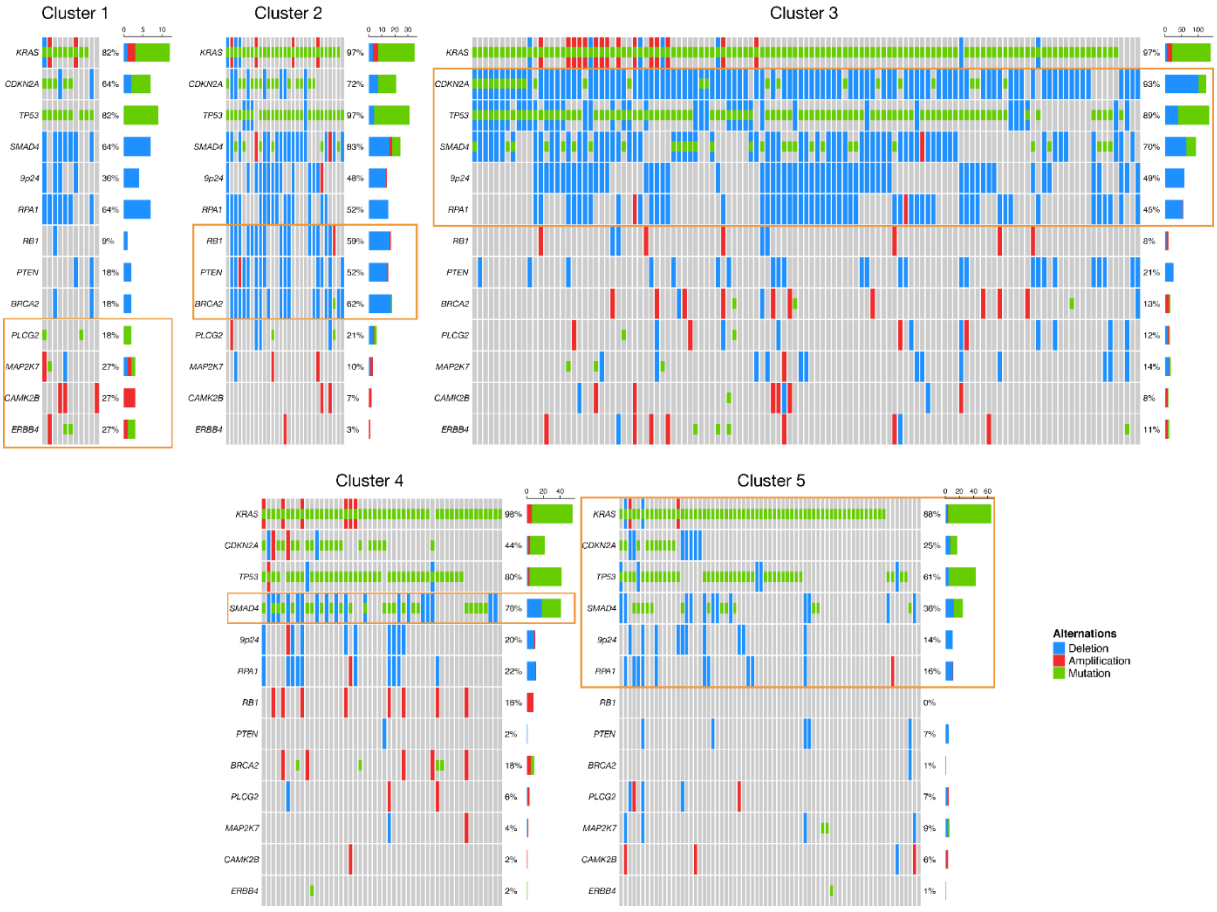


Figure 3.30: Genetic alteration pattern in NMF clusters. Exemplary genetic alterations that experienced significant enrichment in the different NMF clusters. Significantly enriched/depleted genes are circled in orange. Modified from Hoyer et al.[1]

Within single nucleotide mutations, specific base-change patterns have been associated with distinct mutational processes and can be used for targeted therapy selection[175]. To see whether a prevalent, actionable signature existed within one NMF cluster, I analysed the base

changes of all mutations. Two main patterns were observed. First, patients in cluster 1 and cluster 2 had a higher prevalence for transitions than patients in clusters 3, 4, and 5 (80% and 72% versus 58%, 53%, and 58%). Second, patients in cluster 1 had a significant enrichment in A>G/T>C base changes (two-sided Fisher's exact p=0.05), while cluster 4 was significantly enriched in C>A/G>T (two-sided Fisher's exact p=0.002), a signature previously associated with smoking[176] (Figure 3.31).

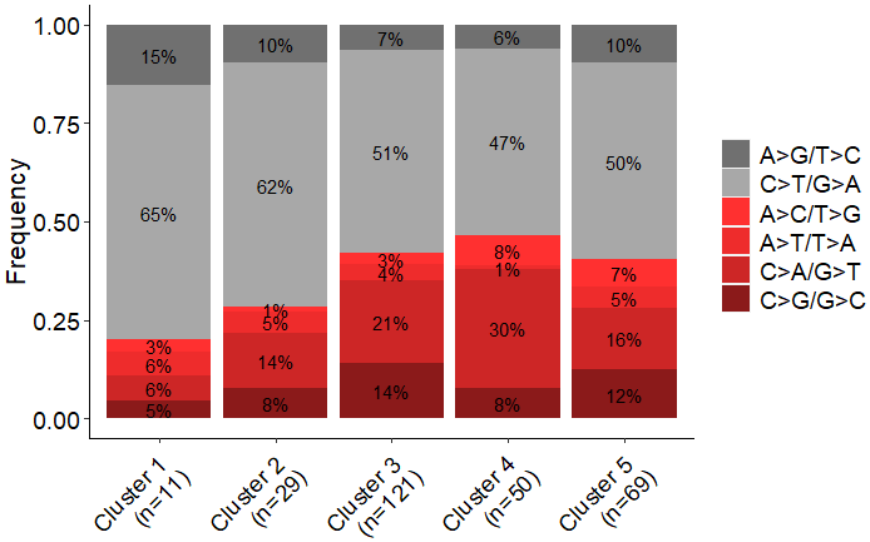


Figure 3.31: Base change signatures for NMF cluster. Overview of shares for all transition (grey) and transversion (red) base changes within each NMF cluster. Modified from Hoyer et al.[1]

3.3.2 Expression pattern in NMF cluster

To assess the cluster-specific genetic patterns' impact on downstream signalling, I compared each cluster's gene expression levels (obtained by Nanostring) with the rest of the cohort (differential expression analysis). Only two clusters showed significant decrease or increase in gene expression. Cluster 2 related significantly to overexpression of genes-encoding receptors and effectors of the PI3K/Akt pathway (*MYB*, *MDM2*). In cluster 5, I noted overexpression of MAPK pathway activating genes (*RASGRP1*, *PDGFA*, and *PRKACB*) as well as the PI3K/Akt inhibitor PTEN, while several cell-cycle control genes had decreased (*PKMYT1*, *SFN*, *CHEK2*, and *SKP2*) (Figure 3.32).

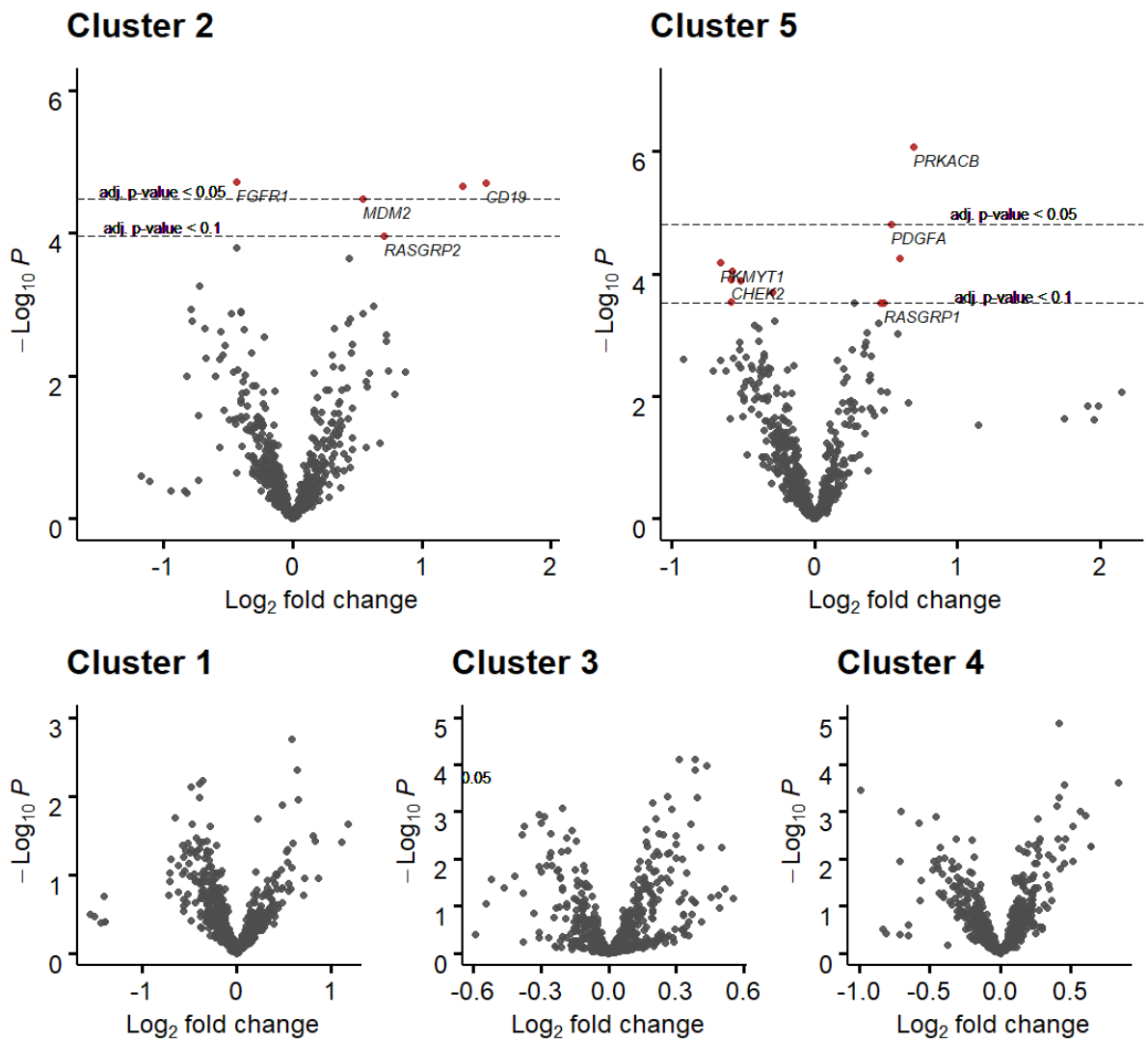


Figure 3.32: Volcano plots showing the results of the differential expression analysis across NMF clusters. The plots depict differential expression analyses, comparing gene expression data (obtained from the Nanostring PanCancerPanel) of each cluster with the rest of the cohort. Horizontal lines show significance levels of p-values (multiple testing adjusted with Benjamini-Hochberg). Variants with an adjusted p-value of < 0.1 are highlighted in red. Genes that are part of the MAPK pathway or the PIK3-/Akt pathway are labelled. Previously published in Hoyer et al.[1]

3.3.3 Survival analysis and metastasis pattern in NMF cluster

I used the Kaplan-Meier survival analysis to compare the DFS and OS of all patients in one cluster against survival from the remaining patients. This revealed two clusters with opposing survival outlooks. Patients in cluster 2 relapsed and died significantly earlier, with a median DFS of 7.3 months and median OS of 15.8 months. Meanwhile, patients in cluster 5 lived significantly longer than the rest of the cohort, with both median DFS and OS double that of cluster 2 (15.6 month and 30.6 month, respectively; Figure 3.33).

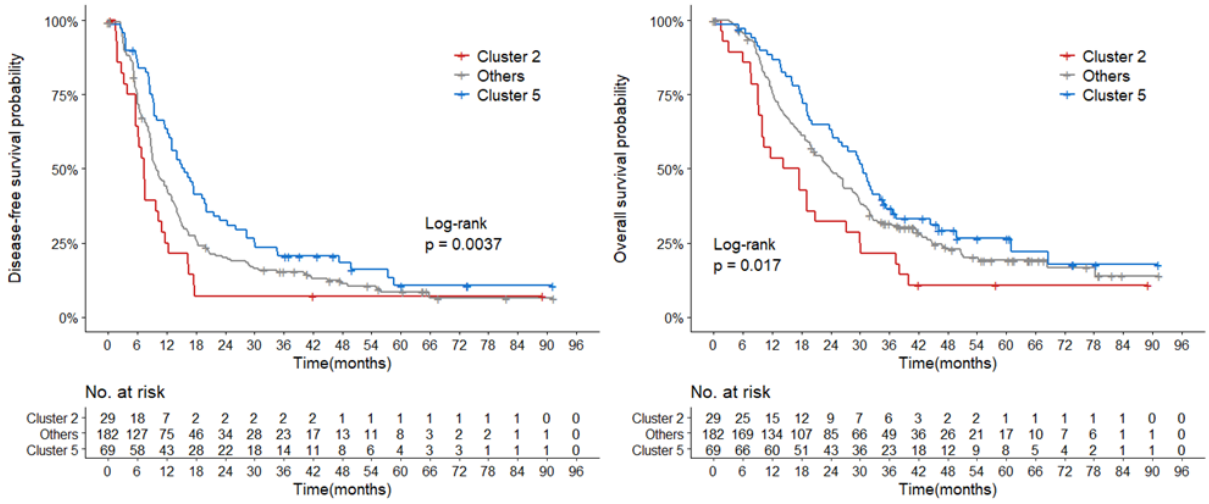


Figure 3.33: Kaplan-Meier curves comparing DFS and OS of cluster 2 and cluster 5 with the remaining patients. Both DFS and OS are significantly decreased for patients in cluster 2 (red) and increased for cluster 5 patients (blue). The group 'Others' includes all patients from clusters 1, 3, and 4. Modified from Hoyer et al.[1]

This effect was independent of other clinical variables. Patients in cluster 2 had significantly shorter DFS and OS rates in multivariate cox regression analyses (HR=1.96, p=0.002 and HR=2.06, p=0.001). In contrast, cluster 5 patients showed longer DFS and OS (HR=0.6, p=0.002 and HR=0.65, p=0.015; Figure 3.34).

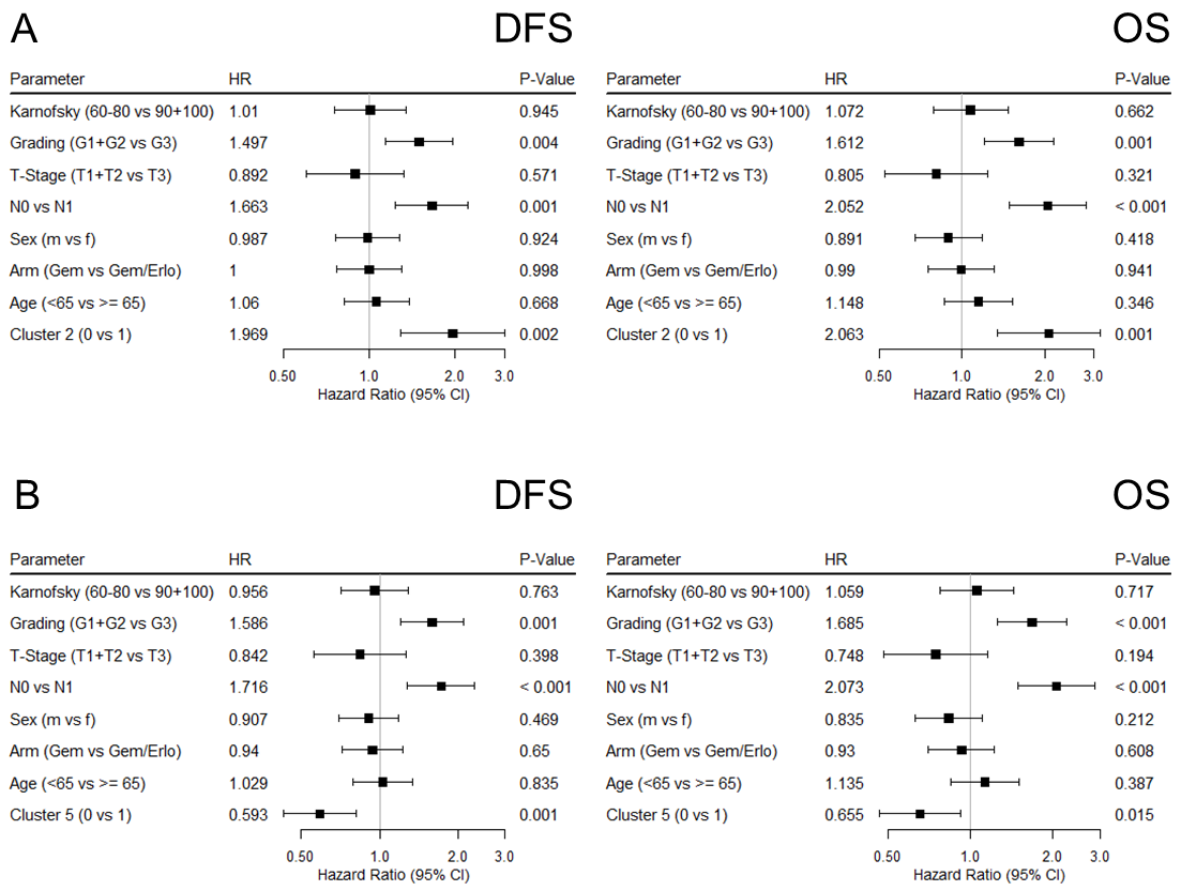


Figure 3.34: Multivariate cox regression analysis for the two NMF clusters with significant differences in DFS and OS. All clinical baseline variables are included in the analysis, which shows a significant independent impact on DFS and OS for **A**) Cluster 2 and **B)** Cluster 5. Modified from Hoyer et al.[1]

The next step was to identify reasons for observed survival differences. To accomplish this, cluster-specific relapse patterns were dissected. Although cluster 2 and cluster 5 showed similar relapse rates (79% vs 74%, respectively; Figure 3.35 A), distributions of relapse occurrence differed. Patients in cluster 2 trended toward more frequent distant metastasis (14% local vs 66% distant) than cluster 5 (23% local vs 52% distant; Figure 3.35 B). Within the distant metastasis, cluster 2 had more patients with metastasis only in the liver (liver only), while patients in cluster 5 tended to have more metastasis only in the lung (lung only), a clinical pattern associated with increased OS in PDAC[171] (Figure 3.35 C). It should be noted that, due to several subdivisions, patient numbers became relatively low – especially in the analysis of distant metastasis patterns – and findings did not reach significance..

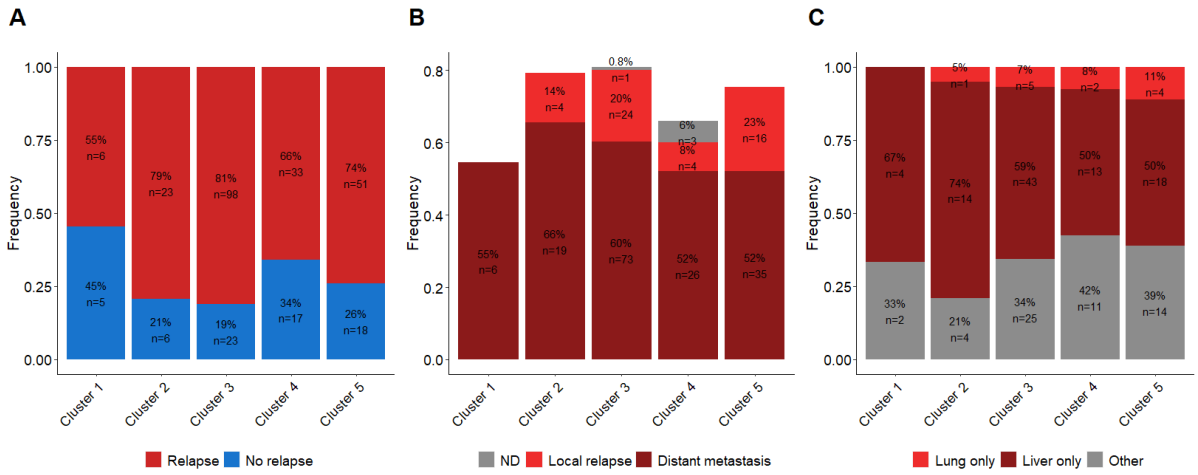


Figure 3.35: Relapse patterns in NMF clusters. A) Proportions of patients with or without relapse in each patient cluster. Total patient number and percentages of total patient number per cluster are depicted. **B)** Further subdivision of the relapse cases into either local or distant metastasis. This depicts total patient number and percentages of total patient number per cluster. **C)** Shares of the different types of distant metastasis, dividing them into lung only (metastasis found only in the lung), liver only (metastasis found only in the liver) and other (other distant metastasis patterns, e.g., lung and liver, peritoneal metastasis). This depicts total patient number and percentages of number of patients with distant metastasis per cluster are.

Comparing DFS and OS between both treatment arms for each cluster separately allowed me to assess whether one of the clusters contained patients who might better benefit from the additive erlotinib. The only cluster to display any increased sensitivity to erlotinib was cluster 4. Here, the patients showed a trend toward longer OS when treated with erlotinib than the gemcitabine-only treatment arm (OS: log rank p=0.089, DFS: long rank p=0.098; Figure 3.36 and Supplemental Figure 7).

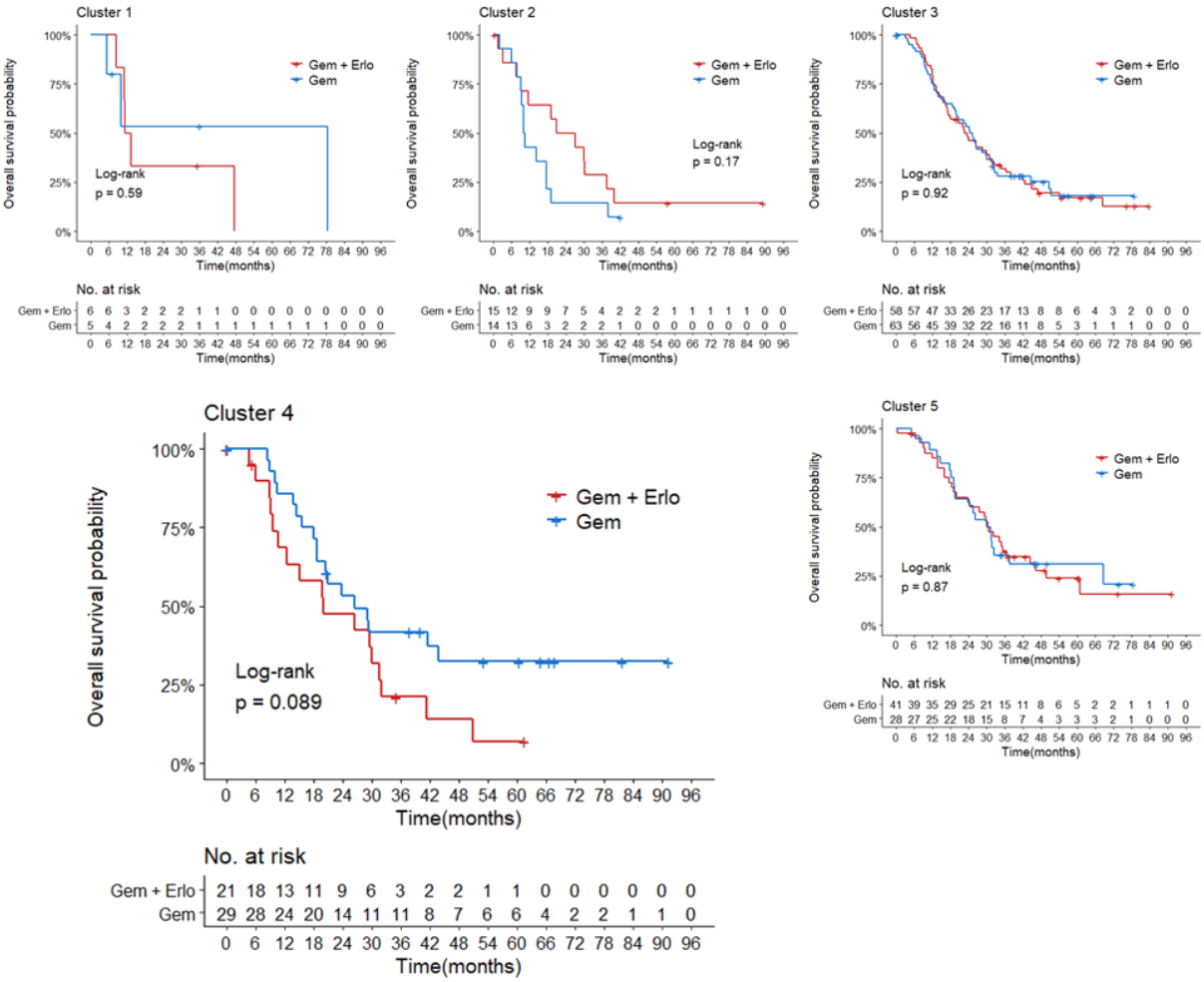


Figure 3.36: Kaplan-Meier curves of OS comparing both treatment arms for each cluster. All patients in each cluster are included and divided depending on their treatment arm. Cluster 4 is the only cluster with a difference in survival based on treatment. Abbreviation: Gem = gemcitabine, Gem + Erlo = gemcitabine and erlotinib combination therapy. Modified from Hoyer et al.[1]

The largest cluster, cluster 3, showed no unique biological or clinical associations. This is likely due to a persisting heterogeneity within this cluster which requires further subgrouping based on even larger patient cohorts. Nevertheless, the NMF clustering did identifiy two well-defined patient clusters (cluster 2 and cluster 5) with distinctive clinical outcome and potentially actionable genetic lesions as well one cluster with a trend for increased erlotinib sensitivity (cluster 4).

3.4 The role of *SMAD4* alterations in clinical outcome and erlotinib sensitivity

One major aim of this thesis was to detect a biomarker for erlotinib sensitivity. Since EGFR is the direct target of erlotinib, alterations in its genetic background represented the most intuitive starting point. As previously seen in advanced PDAC,[123] no predictive or prognostic effect on DFS or OS was observed with respect to the *EGFR* mutation or the CNA status in the CO-2016 cohort (Supplemental Figure 8). However, another genetic alteration did appear to impact erlotinib sensitivity. In NMF cluster 4, which was enriched for *SMAD4* mutations, I noted a trend toward longer OS and DFS in the erlotinib treatment arm. Most *SMAD4* SNVs were truncating mutations and almost all *SMAD4* CNAs were deletions, which suggests similar consequences for loss of function. As a result, patients harbouring either one of these were grouped in a *SMAD4* altered subgroup (*SMAD4*^{alt}, n=179). While *SMAD4*^{alt} status and treatment arm alone were not prognostic for OS in a multivariate cox hazard analysis (Figure 3.37 A), the interaction test for both parameters correlated with a significantly longer OS and DFS (HR=0.53, p=0.033 and HR=0.57, p=0.041 respectively; Figure 3.37 B). Additionally, after correcting for the effect of treatment arms, *SMAD4* alteration status itself became a negative prognostic marker for OS and DFS (OS: HR=1.67, p=0.014; DFS: HR=1.59, p=0.016 respectively Figure 3.37 B, '*SMAD4* [WT vs alt]').

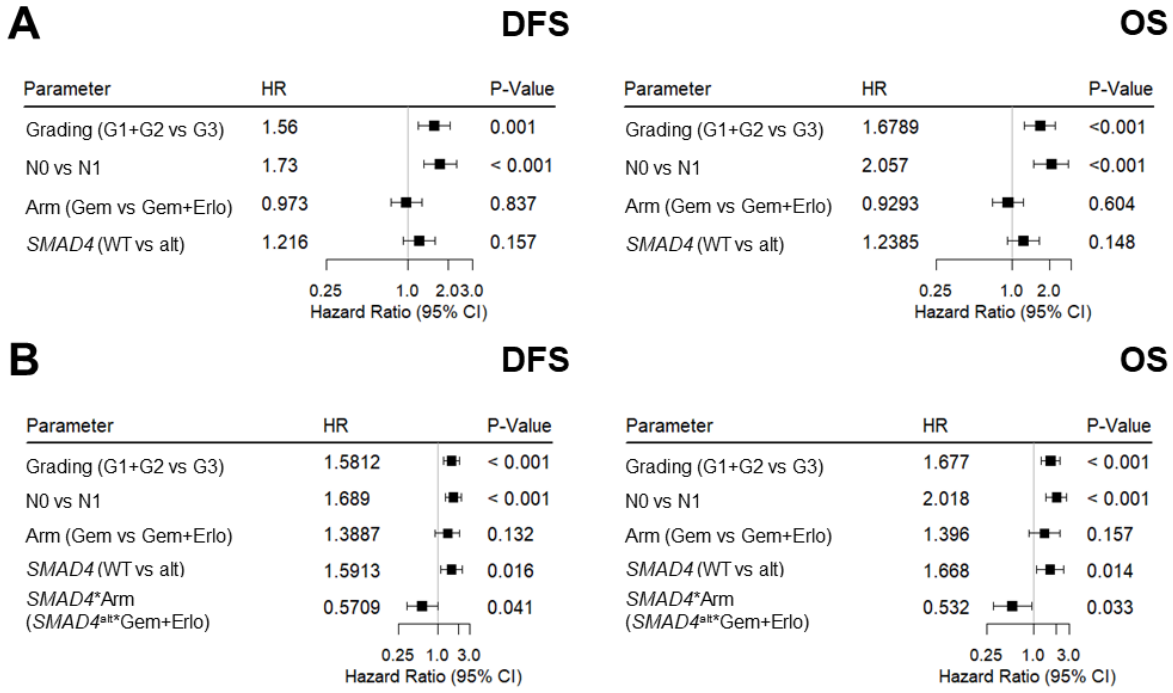


Figure 3.37: Multivariate cox regression analysis of *SMAD4*^{alt} patients and treatment arm. Forest plot of DFS and OS for **A)** both treatment arm and *SMAD4* status as independent variables, which show no significant impact on survival, while **B)** the interaction test of both parameters, shows a significantly decreased HR for OS and DFS in the erlotinib treatment arm of *SMAD4*^{alt} but not *SMAD4* WT patients. Modified from Hoyer et al.[1]

In the gemcitabine treatment arm (Gem), *SMAD4*^{alt} patients had a significantly shorter DFS and OS than *SMAD4* WT patients (log rank: DFS: p = 0.018, OS: p = 0.0078). Simultaneously, *SMAD4*^{alt} patients showed a strong trend toward superior OS when treated with gemcitabine and erlotinib (Gem+Erlo) than gemcitabine alone without reaching significance (log rank: DFS: p = 0.209, OS: p = 0.056). This difference in OS affected approximately 30% of patients who were long-term survivors, leading to a split in the survival curves and subsequently to a clinical course for *SMAD4*^{alt} erlotinib-treated (*SMAD4*^{alt} Gem+Erlo) patients which is comparable to *SMAD4* WT patients (Figure 3.38). Taken collectively, then, *SMAD4*^{alt} patients had a significantly shorter OS and DFS when treated with gemcitabine alone, an effect that was negated when erlotinib was added to the adjuvant chemotherapy regimen.

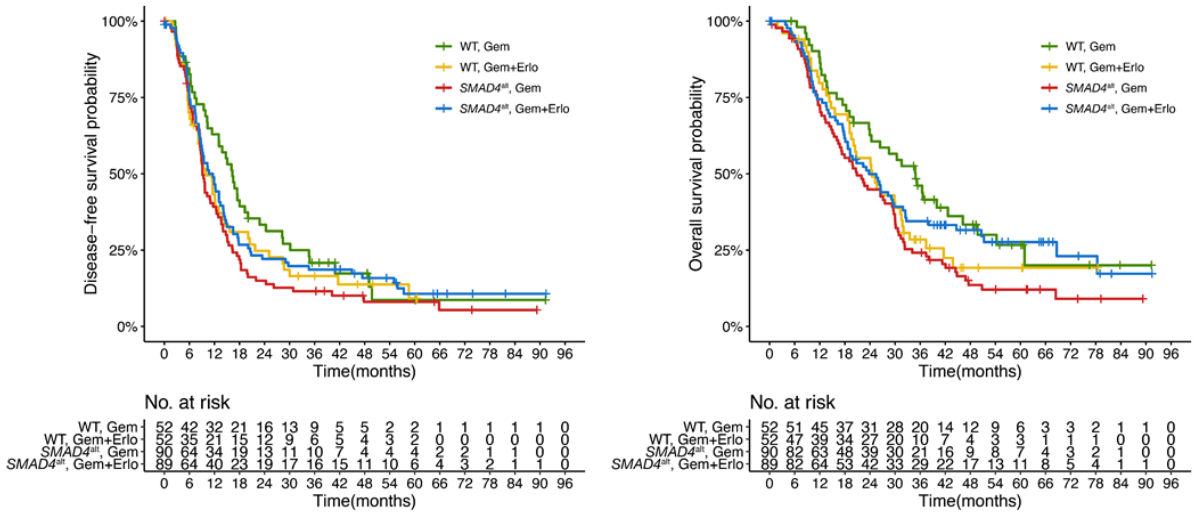


Figure 3.38: Erlotinib sensitivity in *SMAD4*^{alt} PDAC patients. Kaplan-Meier curve comparing OS between both treatment arms in *SMAD4*^{alt} and *SMAD4* WT patients. Modified from Hoyer et al.[1]

To gain insight into differentially activated pathways with respect to the underlying *SMAD4* status, I compared the gene expression profiles of 153 *SMAD4*^{alt} and 69 *SMAD4* WT PDAC patients. After correcting for multiple testing, a total of 11 and 19 genes were significantly up- or down-regulated in *SMAD4*^{alt} patients (Figure 3.39 A, Supplemental Table 9). In line with its frequent genetic loss, *SMAD4* expression itself decreased by 30% in *SMAD4*^{alt} patients. Other down-regulated genes that potentially played important mediating roles were Jun-kinase *MAPK9*, the apoptosis activator *BAD,* and the gene-encoding ubiquitin *UBB*. In contrast, transcription factors *CREBBP* and *MECOM* and tyrosine kinase *KIT* were upregulated in patients with impaired *SMAD4*. The altered genes belong to a complex network of functional interactions that was created using the Reactome Functional Interaction (FI) Cytoscape PlugIn. It links decreased *SMAD4* to both *EGFR* via *UBB* and *MAPK9* as well as the transcriptional activator *CREBBP* and the NFkB pathway (Figure 3.39 B).

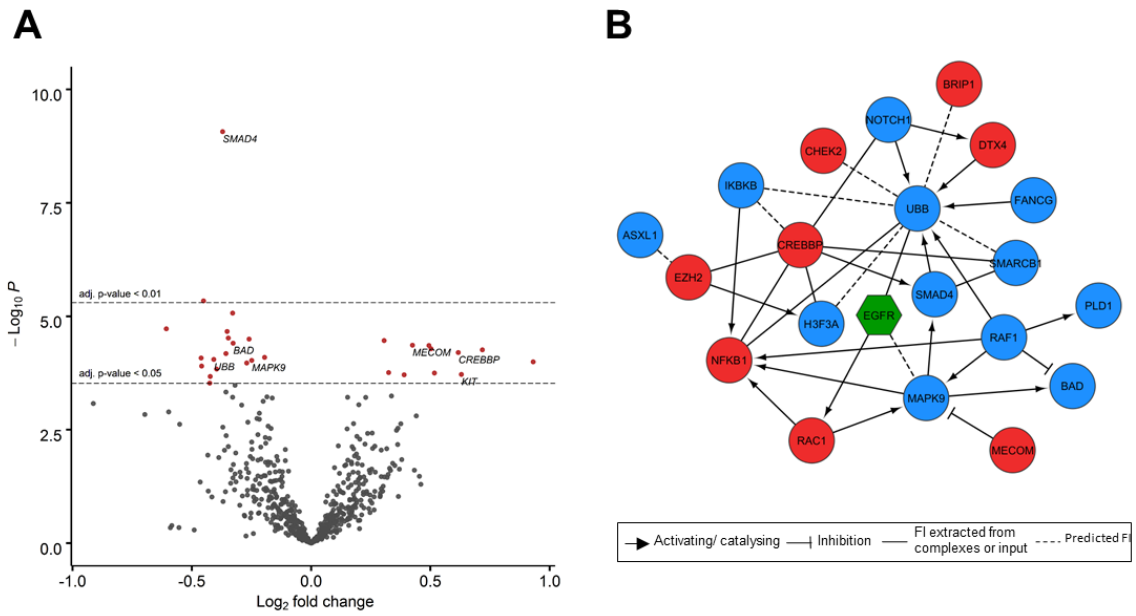


Figure 3.39: Expression profile of *SMAD4*^{alt} patients. A) Volcano plots showing the results of the differential expression analyses between *SMAD4*^{alt} WT patients. Gene expression data (obtained from the Nanostring PanCancerPanel) of each cluster were compared with the rest of the cohort. Horizontal lines show significance levels of p-values (multiple testing adjusted with Benjamini-Hochberg). Variants with an adjusted p-value of <0.05 are highlighted in red. Important signalling genes are labelled. **B)** The functional interaction network across all significant differentially expressed genes (p-value of <0.05) created with the Cytoscape Reactome FI PlugIn. Genes with no direct interactions are not shown. Blue dots are genes with decreased expression in *SMAD4* altered patients and red dots are those with increased expression. The EGFR is included to model possible interactions with the erlotinib effector and is not differentially expressed in either group. Interaction types are depicted by different connecting lines. Abbreviations: FI = functional interaction. Modified from Hoyer et al.[1]

Increased JNK activation has shown to mediate acquired resistance to EGFR inhibition by bypass signaling[177]. Therefore, I decided to add greater detail to my analysis of the effect of *MAPK9* expression on *SMAD4* mediated erlotinib sensitivity. Integrating *SMAD4* genetic aberration status with differential gene expression data grouped *SMAD4*^{alt} patients into patients with low (91/222=40.9 %) or high (62/222=27.9 %) *MAPK9* expression levels (Table 3.6).

Table 3.6: Number of patients in each *SMAD4*^{alt} *MAPK9* subgroup. Patients were divided by *SMAD4* status and *MAPK9* expression levels. *MAPK9* high and low was calculated through median *MAPK9* expression in all 230 patients with available expression values. Due to missing information about *SMAD4* status, only 222 patients were included in the final analysis. Previously published in Hoyer et al.[1]

	SMAD4^{alt}	*SMAD4*^{WT}	**total**
***MAPK9*^{low}**	91	22	113
***MAPK9*^{high}**	62	47	109
total	153	69	

Strikingly, the beneficial effect of erlotinib was restricted to *SMAD4*^{alt} patients with low *MAPK9* expressions (DFS: HR=0.49; test for interaction, p=0.02) and OSes (HR=0.32; test for interaction, p=0.001; Figure 3.40).

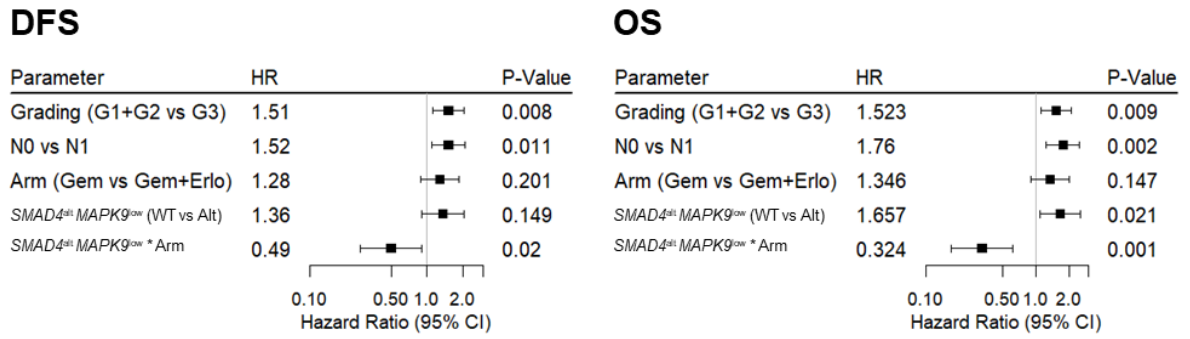


Figure 3.40: Multivariate cox regression analysis of *SMAD4*^{alt} *MAPK9*^{low} patients and treatment arm. Forest plot of DFS and OS for interaction of both treatment arms and *SMAD4 MAPK9* status show a significant impact on survival. Modified from Hoyer et al.[1]

As a result, when treated with erlotinib, *SMAD4*^{alt} MAPK9^{low} patients lived almost twice as long as patients without erlotinib (median OS: 29.7 vs 16.0 months) and had a median of three more months before relapsing (median DFS: 11.9 vs 9.0 months). This effect was very specific, and no survival increases upon erlotinib treatment were observed in either *SMAD4*^{alt} patients with high *MAPK9* expression nor in *SMAD4*^{WT} patients with low *MAPK9* expression levels (Figure 3.41, Supplemental Figure 9).

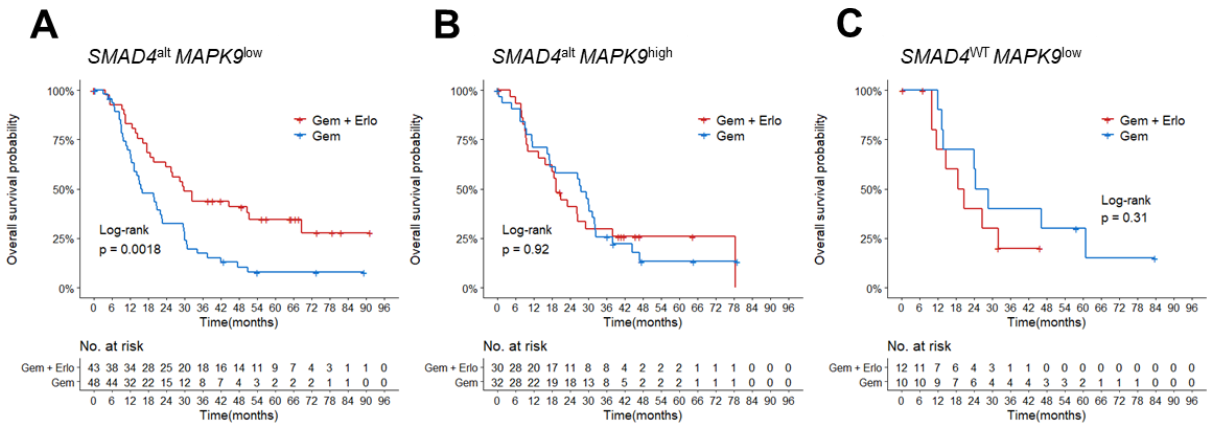


Figure 3.41: Kaplan-Meier curves of treatment arms within patient groups with different *SMAD4 MAPK9* statuses. A) *SMAD4*^{alt} *MAPK9*^{low} patients lived significantly longer when treated with erlotinib, while neither **B)** *SMAD4*^{alt} *MAPK9*^{high}, nor **C)** *SMAD4*^{WT} *MAPK9*^{low} patients exhibited that effect. Modified from Hoyer et al.[1]

With most of the patients relapsing eventually, both time to relapse as well as specific relapse patterns can significantly impact long-term survival and the course of disease. Loss of *SMAD4* has been associated with pancreatic tumour progression and increased metastasis formation in PDAC[178,179]. When examining the data on relapse patterns within the 222 patients with

known *SMAD4* *MAPK9* status, two trends were observed. First, the percentage of patients without relapse in erlotinib treated *SMAD4*^{alt} *MAPK9*^{low} patients was nearly double that of the patients treated with gemcitabine alone (19% versus 10%, respectively), while no such change occurred in the remaining patients following erlotinib treatment (Figure 3.42 A). In addition, a trend toward distant, lung-only metastasis was noted in *SMAD4*^{alt} *MAPK9*^{low} patients following erlotinib treatment. Their proportion was four times as high as in the rest of the cohort (12% versus 3%), while the same ratio was nearly 1:1 in the gemcitabine treatment arm (7% versus 8%; Figure 3.42 C). This pattern has previously been associated with longer OS in PDAC[171] and explain the increased OS in *SMAD4*^{alt} *MAPK9*^{low} patients treated with erlotinib. It should be noted that, due to several subdivisions, patients numbers became relatively low – especially in the analysis of distant metastasis patterns. Therefore, the findings did not reach significance and need to be further validated.

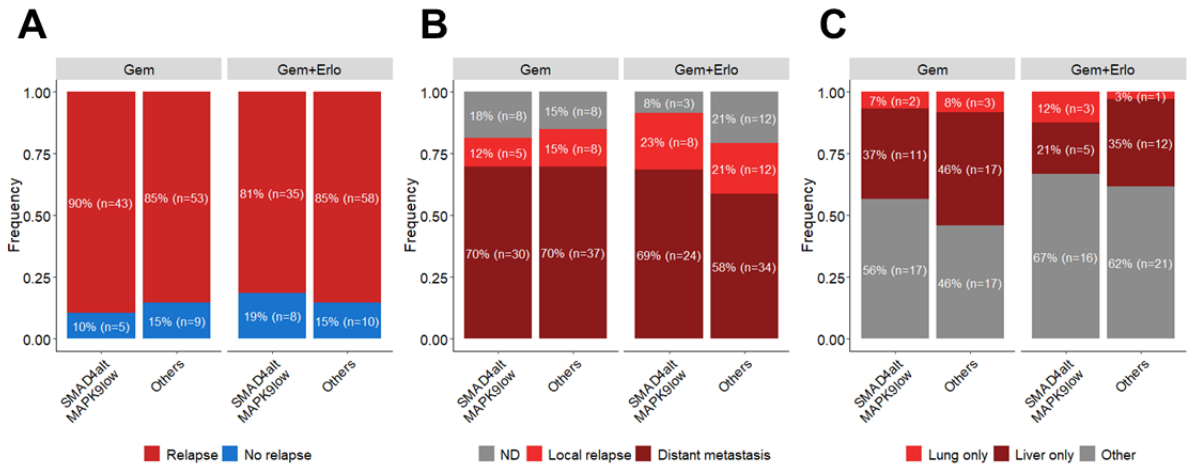


Figure 3.42: Relapse pattern in *SMAD4*^{alt} *MAPK9*^{low} patients divided by treatment arm. A) Proportions of patients with or without relapse for *SMAD4*^{alt} *MAPK9*^{low} patients compared to the rest of the cohort. Total patient numbers and percentages of total patient number are depicted. **B)** Further subdivision of the relapse cases into either local or distant metastasis. Total patient numbers and percentages of patient numbers with relapse are depicted. **C)** Proportions of the different types of distant metastasis, dividing them into lung only (metastasis found only in the lung), liver only (metastasis found only in the liver), and other (other distant metastasis patterns, e.g., lung and liver, peritoneal metastasis). Total patient numbers and percentages of patient numbers with distant metastasis are depicted.

3.5 Modelling of *SMAD4*^{alt} *MAPK9*^{low} status in PDAC cell line

As a first step toward understanding the effects of SMAD4 and MAPK9 on erlotinib sensitivity in the CO-2016 cohort, I created an *in vitro* model in a PDAC cell line. Choosing the best PDAC cell line, creating several stable clonal CRSPR/cas9 knockouts (KO) of *SMAD4,* and decreasing the *MAPK9* mRNA levels via siRNA knock down allowed me to create a reliable in vitro model to analyse the effect of erlotinib treatment.

3.5.1 Characterisation of five PDAC cell lines as basis for the *in vitro* model

I began by analysing five PDAC cell lines with different genetic backgrounds to find a fitting cell line to modify. Sequencing the five cell lines with the same PDAC specific NGS panel used for the CO-2016 cohort revealed a broad spectrum of SNVs with mutation frequencies comparable to the patients. Two cell lines had nonsense mutations in the *SMAD4* gene (HPAC, Capan1), and one cell line did not show any reads covering the *SMAD4* coding sequence, suggesting bi-allelic loss (BXPC3; Table 3.7).

Table 3.7: SNVs in PDAC cell lines. A list of all SNVs found within the 67 genes included in the PDAC panel for all five cell lines. Variants that could not be found within the COSMIC cell lines database are depicted in orange.

	PANC1	HPAC	BXPC3	CAPAN1	CAPAN2
KRAS	G12N	G12D	-	G12V	G12V
TP53	R273H	G187R	Y220C	A159V	-
SMAD4	-	K49fs	-	S343*	-
CDKN2A	-	E120*	-	-	A18fs
GNAS	A467S	-	-	R393S	-
ARID1A	-	P1631fs	-	-	-
CAMK2B	-	V515M	-	-	-
PREX2	-	I517V	-	-	-
ACVR1B	-	I141L	-	-	-
PRKCG	-	G283S	-	-	-
KDM6A	-	splice	-	-	-
BRAF	-	-	V486fs	-	-
PIK3R5	-	-	A203T	-	-
RNF43	-	-	S495Y	-	R329fs
PIK3CG	-	-	-	T827R	-
ATM	-	-	-	R1585S	T2640I
BRCA2	-	-	-	S1981fs	-
MAP2K4	-	-	-	E232*	-

Detection of the underlying CNA with the MLPA assay confirmed an explanation for the missing *SMAD4* reads in BXPC3. The cell line carries a homodeletion of *SMAD4* and a homozygous deletion of *CDKN2A*. The absence of both copies of *CDKN2A* was a common pattern seen also in PANC1 and CAPAN1. Along with the nonsense mutations in HAPC and CAPAN2, no

cell line carried a fully functioning *CDKN2A* genotype. *SMAD4,* however, were observed as WT in two cell lines (PANC1, Capan2; Table 3.8).

Table 3.8: CNAs in PDAC cell lines. A list of all CNAs found within the 11 MLPA genes for all five cell lines. Deletions are highlighted in blue and amplifications in red. Additionally indicated is whether only one (hetero) or both alleles (homo) were affected.

	PANC1	HPAC	BXPC3	CAPAN1	CAPAN2
APC				hetero	
CDKN2A	homo		homo	homo	
BRCA1					
RB1					
WT1					
SMAD4			homo	hetero	

To study the effects of the *SMAD4* knock-out, the unmodified *cell* line needed be capable of producing full-length protein. Current knowledge of the cell line genotypes allowed for estimation of *SMAD4* expression levels. Indeed, BXPC3 and its *SMAD4* homozygous deletion did not express any *SMAD4* mRNA. In the two cell lines with *SMAD4* SNVs (HPAC, CAPAN1), decreased levels of *SMAD4* could still be detected. CAPAN2 had a *SMAD4* expression comparable to HPAC, while PANC1 had the highest *SMAD4* expression levels (Figure 3.43 A). The effects of the SNVs on *SMAD4* in HPAC and CAPAN1 could subsequently be seen on the protein level where no SMAD4 protein was detected (Figure 3.43 B). The same was true for BXPC3 and CAPAN2. While no SMAD4 protein was expected for BXPC3, the absence of the protein in CAPAN2 could not be explained by its genotype. A range of possible post-transcriptional modifications or epigenetic silencing might be responsible for the protein loss and might have left PANC1 as the only cell line producing full-length SMAD4 protein.

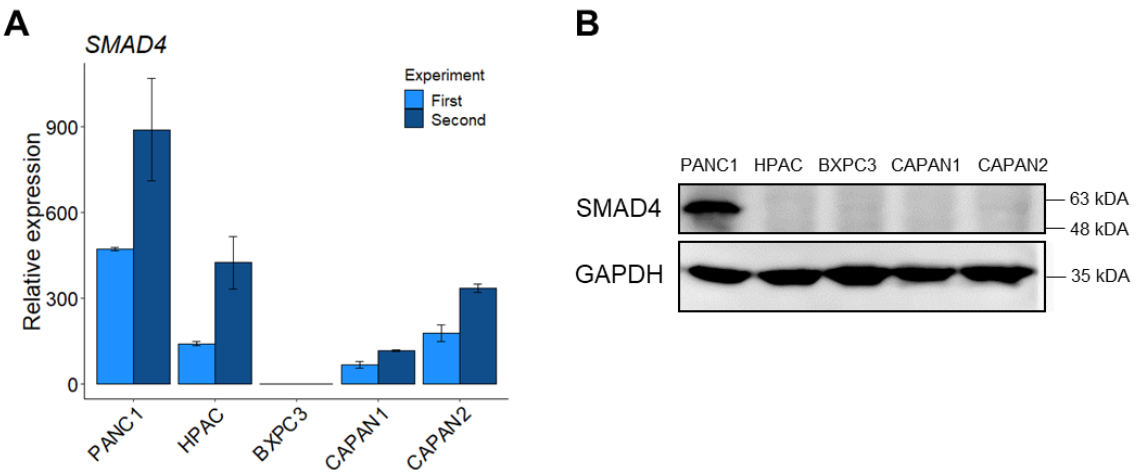


Figure 3.43: SMAD4 mRNA and protein levels in the five cells lines. A) ΔΔCT values normalised to GAPDH BXPC3 were used as reference. The qPCR measurement was repeated twice with RNA extracted at different points in time and each condition was measured in triplicates. **B)** Western blot with antibody against SMAD4 was used as loading control. GAPDH served as loading control.

The two treatment arms of the CONKO-005 trial were gemcitabine and gemcitabine plus erlotinib treatment. To model this treatment *in vitro*, I used a cell proliferation assay based on WST-1 to measure the number of cells still alive after 96 hours of treatment with serial dilutions of the two chemicals. First, the cells were treated with either substance alone. CAPAN1 and BXPC3 were most sensitive to gemcitabine with an IC_{50} of 12 and 38 nM. The IC_{50} of HPAC was comparable (15nM), but its final cell count was much higher, leaving almost 50% of cells resistant to gemcitabine treatment. CAPAN2 sensitivity was already much lower with an IC_{50} of 354nM. PANC1 had an IC_{50} of 223 µM, which is 20,000 times higher than that of CAPAN1. The strong gemcitabine resistance to PANC1 has been previously described[180]. Erlotinib treatment showed an effect on HPAC, CAPAN1 and CAPAN2, with different final cell viabilities and IC_{50} values (9.3 µM, 26.7 µM and 1.2 µM, respectively). In contrast, neither BXPC3 nor PANC1 exhibited a decrease in cell viability after erlotinib treatment (Figure 3.44 A). To mimic the combination treatment from the clinical trial, I combined the IC_{50} concentration of gemcitabine and diluted erlotinib with a serial dilution. In these results, HPAC, Capan1, and Capan2 remained sensitive to erlotinib, while PANC1 and BxPC3 did not show any cytotoxic reaction after added erlotinib treatment (Figure 3.44 B).

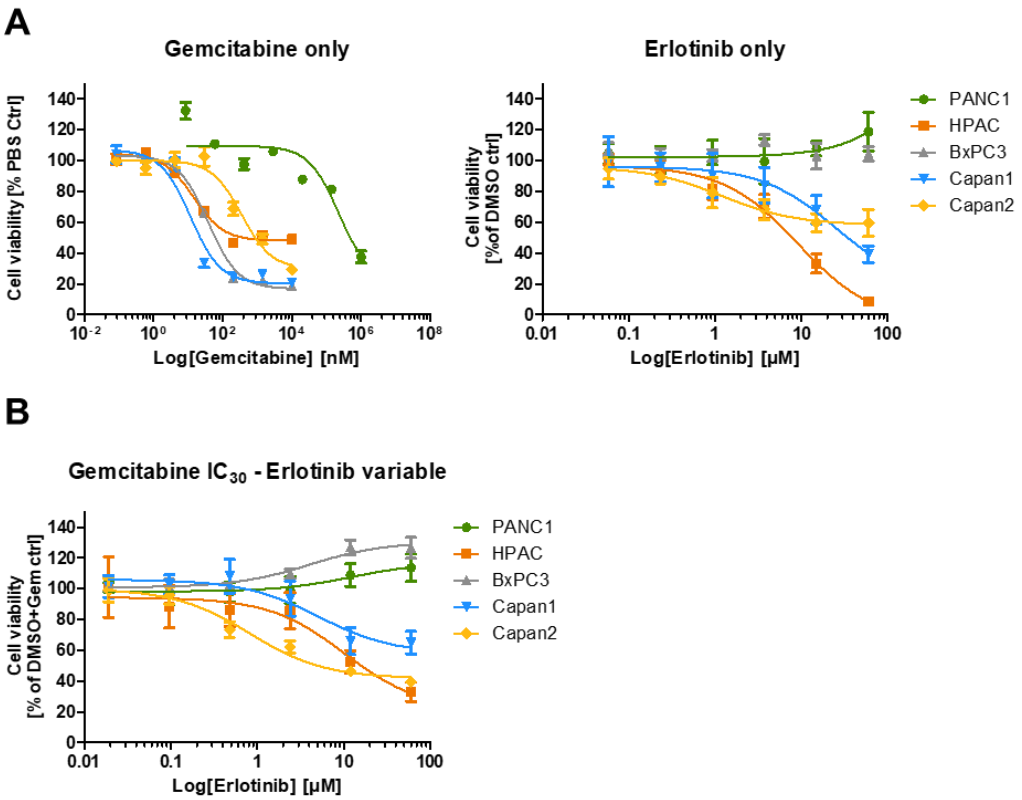


Figure 3.44: Chemosensitivity of unmodified PDAC cell lines. For each cell line, a dose-response curve is fitted through normalized WST absorption values at different gemcitabine and erlotinib concentrations. **A)** Fit of single-agent dose response shows the reaction of the five cell lines to either gemcitabine or erlotinib. **B)** Dose response curves of a combination of gemcitabine and erlotinib. The agents were a stable concentration of gemcitabine comprising of the IC_{50} concentration from A, which was added along with changing concentrations of erlotinib (Gemcitabine IC_{50} – Erlotinib variable).

3.5.2 Generation of *SMAD4* knock-out clones via CRISPR/Cas9

PANC1 revealed a *SMAD4* WT genotype, high *SMAD4* mRNA and protein levels, and a resistance to erlotinib; these factors made it the best cell line for the creation of *SMAD4* KO clones. I began with three single-guide RNAs (sgRNAs) against targets exon 3, exon 6, and exon 9, which I used for a Cas9-sgRNA RNP-based introduction of site-specific indels. One week after transfection, all three sgRNAs had indiced indels with high knockout rates (Figure 3.45).

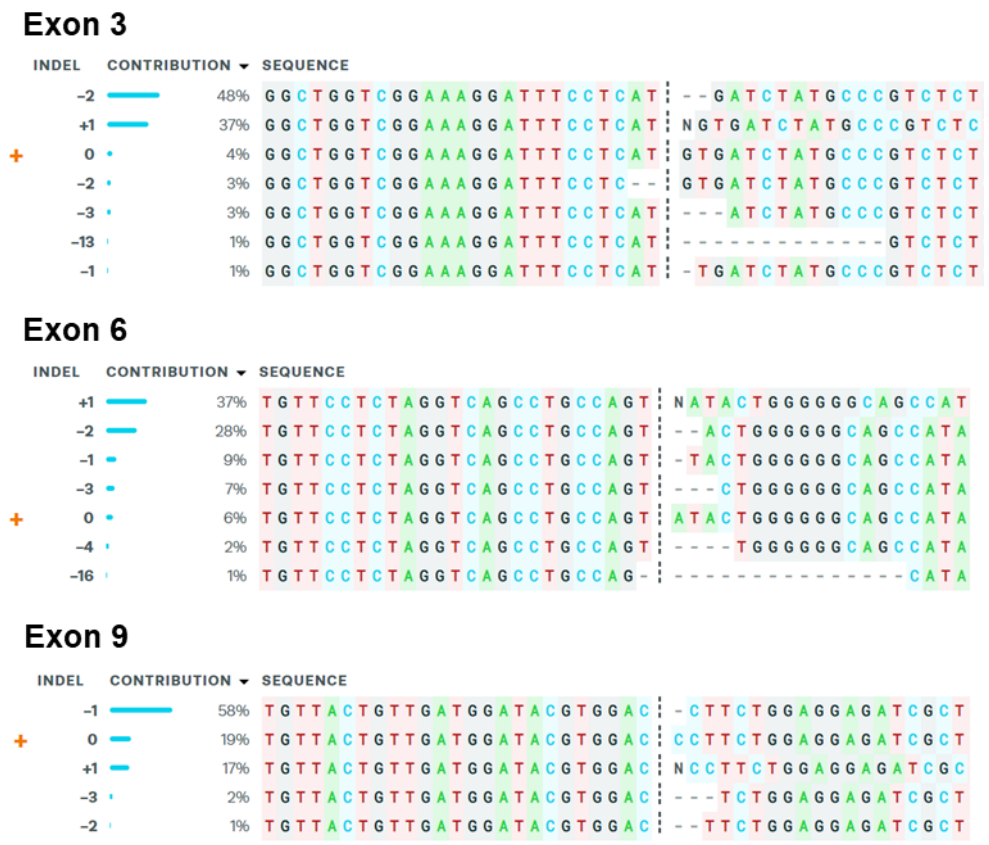


Figure 3.45: Overview of the three CRIPR/Cas9 *SMAD4* KO variants. Deconstructed Sanger sequencing tracks created using ICE analysis to calculate each indel's proportion of the complete Sanger track. Normalized relative contribution of each sequence found in the bulk one week after transfection is shown. Cut sites are represented by black vertical dotted lines, and the wild-type sequence is marked by a '+' symbol on the far left. [Modified from https://ice.synthego.com/[146]].

I subsequently cultured the transfected cell bulk for 37 days to obtain enough cells for single cell cloning and protein extraction. During this time, DNA was extracted and Sanger-sequenced two more times, revealing the development of the mutated population. While the initial knockout rate was remarkably high for all three sgRNAs (78%–93%), the remaining WT cells soon began to overgrow the *SMAD4* mutated cells, leading to up to 92% WT cells (Figure 3.46).

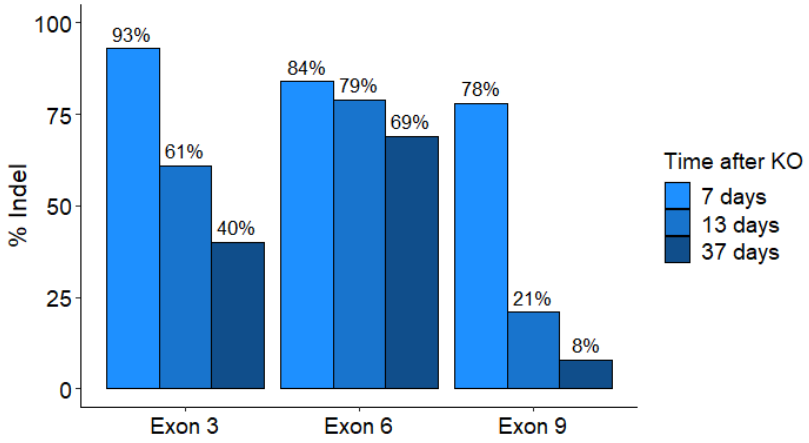


Figure 3.46: Time course of indel frequencies in *SMAD4* KO bulk cells. Share of indels within the cell bulk at different points following initial transfection (Time after KO) show the WT cell overgrowing the relatively persistent knockout cells over the course of longer than a month.

To confirm the effect of the genetic knockout on *SMAD4* levels in the cell bulk, RNA and protein were extracted 16 days after transfection for the two variants with sufficient cells (Exon 3, Exon 6). While *SMAD4* mRNA levels were slightly higher in the KO cells than in the unmodified cells and cells transfected without sgRNA (empty), no SMAD4 protein was detected in either KO variant (Figure 3.47).

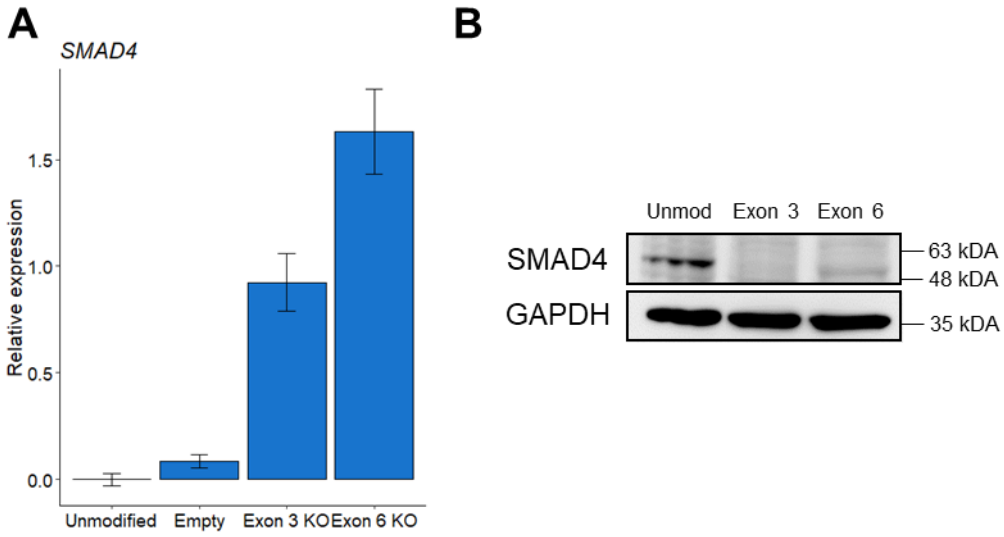


Figure 3.47: *SMAD4* mRNA expression and protein levels in the *SMAD4* KO cells. A) ΔΔCT values normalised based on *GAPDH* CT and ΔCT values of the non-transfected negative control (unmodified). **B)** Western blot with antibody against SMAD4 shows that only the unmodified PANC1 cells express SMAD4. GAPDH served as loading control.

With both KOs confirmed successes, I decided to use the exon 3 KO as basis for single-cell clone generation. The KO occurred this early in the sequence, so the probability that reactive protein fragments would be expressed was less likely. Therefore, the cell bulk was single-cell sorted into two 96-well plates 37 days after transfection and incubated for another 20 days. Single cell colonies were found in 21 wells, which were then further expanded for more than

20 days when colonies had grown fully confluent in a 6-well plate. At this point, DNA was extracted and cells genotyped by sequencing. Of the twelve evaluable, one clone was WT for SMAD4, while the other clones showed varying mutation patterns. Except for two, all clones exhibited only one or two alterations in 100% and 50% of cells, respectively, a sign of clean clonal populations (Supplemental Figure 10).

For subsequent analysis, I chose four clones covering all possible *SMAD4* mutation patterns: the WT clone 2; the homozygous deleted clone 19; the heterozygous clone 20; and clone 17, which has an insertion and a deletion in *SMAD4* (Figure 3.48 A). Comparison of their respective *SMAD4* mRNA expression levels showed a slight increase for all three *SMAD4*-altered clones, while the WT clone 2's *SMAD4* mRNA expression decreased from that of unmodified PANC1 cell bulk (Figure 3.48 B). Western blot analysis in turn showed that the *SMAD4* double KOs (clone 17, clone 19) possessed no detectable SMAD4 protein, while both the WT clone and the heterozygous mutated clone (clone 2, clone 20) still produced SMAD4 (Figure 3.48 C).

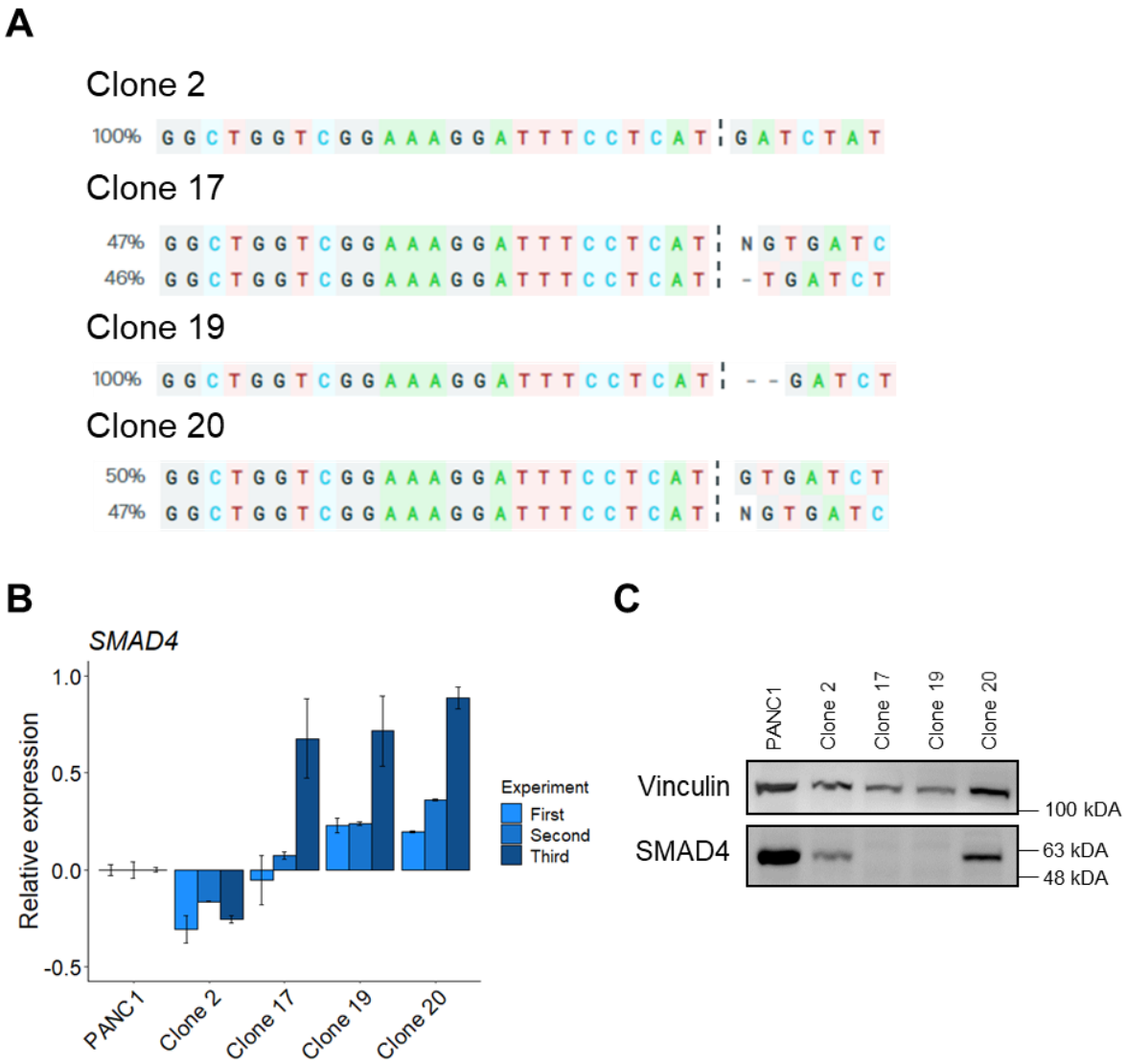


Figure 3.48: Characteristics of selected *SMAD4* KO single cell clones. A) Genotype of the four single-cell clones chosen for further analysis. Clone 2 was WT for *SMAD4,* Clone 20 had a frameshift insertion in one strand, and clones 17 and 19 had both copies of *SMAD4* altered [Modified from https://ice.synthego.com/]. **B)** *SMAD4* expression levels in the original cell lines and the four clones increased in *SMAD4* transcripts in the *SMAD4* KO cells compared to PANC1 and the WT clone. ΔΔCT values were normalized based on *GAPDH* CT and ΔCT values of PANC1. **C)** Western blot with antibody against SMAD4 showed that cell lines with SMAD4 mutations in both strands (clone 17, clone 19) produced no SMAD4 protein, while the *SMAD4* WT and heterozygous clones (clone 2, clone 20) did.

Cell proliferation was not affected by the *SMAD4* KO. Both *SMAD4* WT cells (Clone2 and PANC1) exhibited the same duplication rates as *SMAD4* KO cells (Clone 17, Clone 19 and Clone 20; Figure 3.49, Supplemental Figure 11).

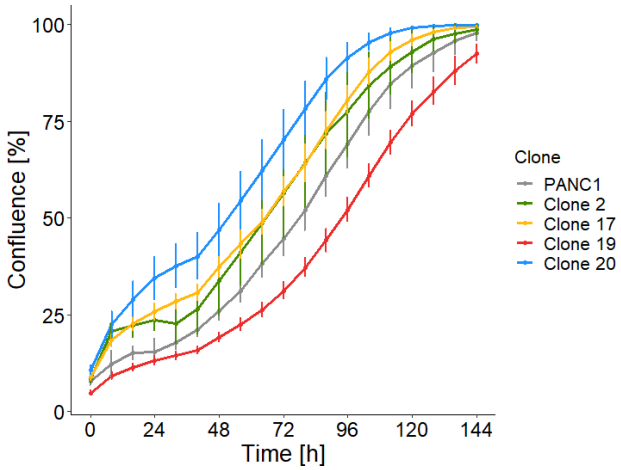


Figure 3.49: Proliferation of single cell clones and PANC1 bulk cells. For cell proliferation measurement 4x10^3 cells were continuously monitored for eight days with the IncuCyte Live-Cell Analysis Systems. Pictures were taken every eight hours and each cell line was measured in quadruplicates. Lines and error bars show mean ± standard deviation of quadruplicates. The experiment was repeated three times (Supplemental Figure 11). No significant difference in cell proliferation was noted between any of the four clones and the cell line.

SMAD4 has shown to affect metastasis patterns *in vivo*. Therefore, I analysed cell migration ability using a wound scratch assay. To simplify the experimental set-up, only the two clones with the most distinct genotypes were included: the WT clone 2 and the homozygous-mutated clone 19. The first observed effect was that *SMAD4*-deficient cells migrated more slowly than their WT counterparts. The second finding was that erlotinib treatment, either alone or combined with gemcitabine, decreased cell migration even further. This effect was present in both clones and therefore appeared to be independent from *SMAD4* status (Figure 3.50, Supplemental Figure 12).

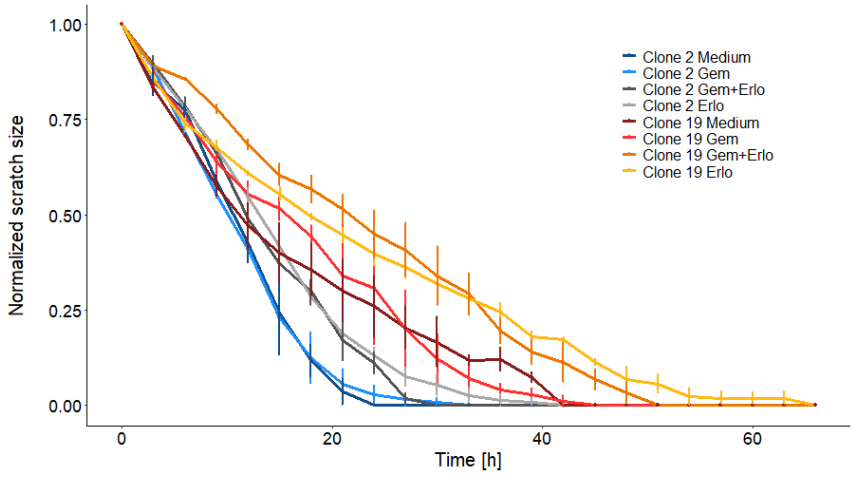


Figure 3.50: Cell migration of *SMAD4* KO single-cell clones with or without treatment. The results of the wound scratch assay comparing the WT clone (clone 2) with the SMAD4 KO clone (clone 19). For each clone 1.5x10^5 cells were incubated in different medium conditions. After applying the scratch, cells were continuously observed for four days. Pictures were taken every three hours in triplicates for each condition. Lines and error bars show mean ± standard deviation of triplicates. The experiment was repeated three times (Supplemental Figure 12). *SMAD4* KO decreased cell migration ability, an effect that is enhanced when treated with erlotinib.

SMAD4 KO had no effect on erlotinib sensitivity; *SMAD4*- cells (clone 17, clone 19, clone 20) showed the same response to increasing erlotinib concentrations as *SMAD4* WT cells (PANC1, clone 2). Interestingly, the gemcitabine sensitivity of the *SMAD4* KO cells was comparable to that of the *SMAD4*- cell lines (HPAC, BXPC3, CAPAN1, CAPAN2; Figure 3.51).

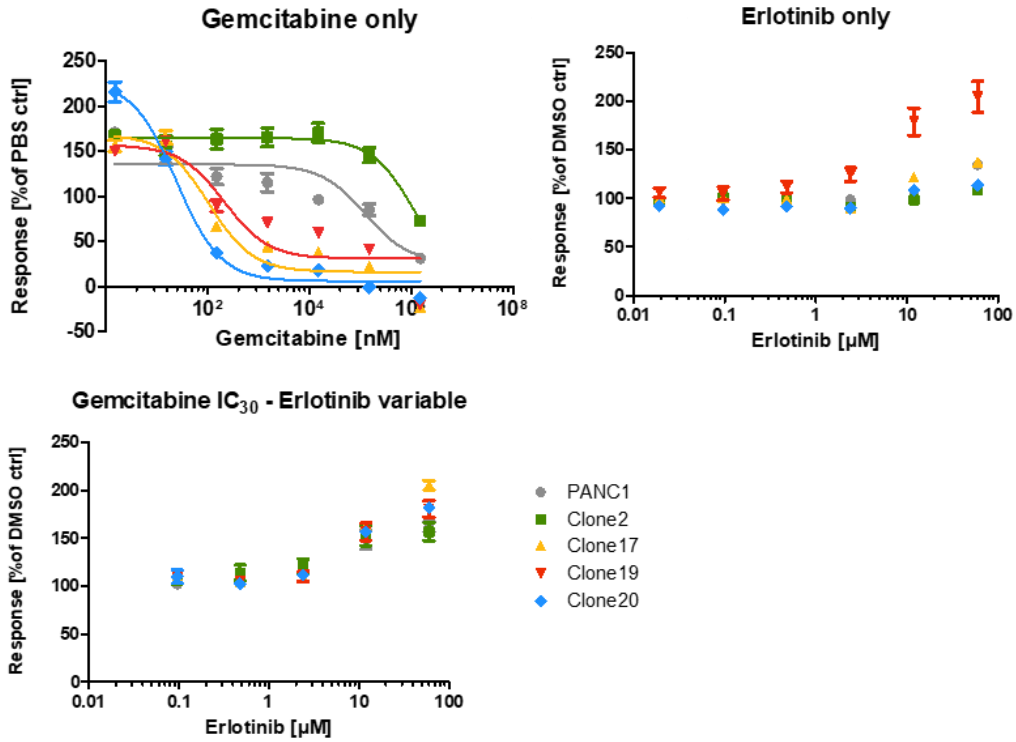


Figure 3.51: Chemosensitivity of *SMAD4* KO SCCs. For each clone, a dose-response curve is fitted through normalized WST absorption values at different gemcitabine and erlotinib concentrations. All three *SMAD4* altered clones (clone 17, clone 19, clone 20) showed increased gemcitabine sensitivities but no changes in erlotinib-based cell toxicity. This effect was observed both with single agents and with the combined treatment of erlotinib and gemcitabine.

To better understand the signalling events responsible for the difference in SMAD4-deficient (*SMAD4*-) cell behaviour, I analysed expression levels of the genes that were found to be differentially expressed between $SMAD4^{alt}$ and $SMAD4^{WT}$ patients. Three genes were upregulated in the *SMAD4*- cell lines: *MAPK9*, *MECOM* and *KIT* (Figure 3.52, Supplemental Figure 13). The expression values were normalized to the unmodified PANC1 cell line. This cell line typically has a medium *MAPK9*, a high *KIT,* and a very low *MECOM* expression when compared to other PDAC cell lines (Supplemental Figure 14).

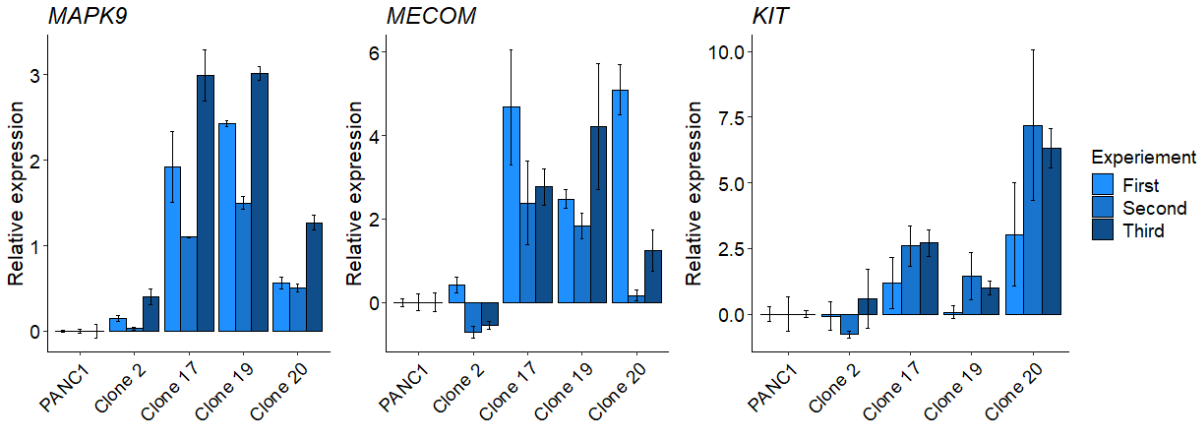


Figure 3.52: Expression of *MAPK9*, *MECOM* and *KIT* in *SMAD4* KO single cell clones. Relative expression of three genes found to be significantly increased or decreased in *SMAD4*^{alt} patients and that had a different expression in SMAD4 KO cells. ΔΔCT values normalized based on *GAPDH* CT and ΔCT values of PANC1.

3.5.3 Temporary siRNA knock down of *MAPK9* in *SMAD4*^{alt} cells

A significant effect on DFS and OS after additive erlotinib treatment in the CO-2016 cohort was seen only in *SMAD4*^{alt} patients with a low *MAPK9* expression. Therefore, to model both parameters accordingly, I used a siRNA against *MAPK9* in the *SMAD4* KO single-cell clones. A siRNA against GFP was used as the treated negative control. The *MAPK9* knockdown worked very well. This was observed on the RNA level where treatment with *MAPK9* siRNA decreased *MAPK9* RNA to 9–35% of the expression in unmodified cells (Figure 3.53 A). *MAPK9* siRNA knockdown also significantly decreased MAPK9 protein in treated cells (Figure 3.53 B). This had no effect on either the SMAD4 or EGFR proteins or mRNA levels in the two clones chosen for subsequent analysis (clone 2, clone 19). C-jun, the downstream effector of MAPK9, showed increased expression and phosphorylation following *MAPK9* siRNA treatment. The same effect was noted in the GFP-treated negative control cells and therefore appeared to be an unspecific response related to stress and siRNA treatment (Figure 3.53 B).

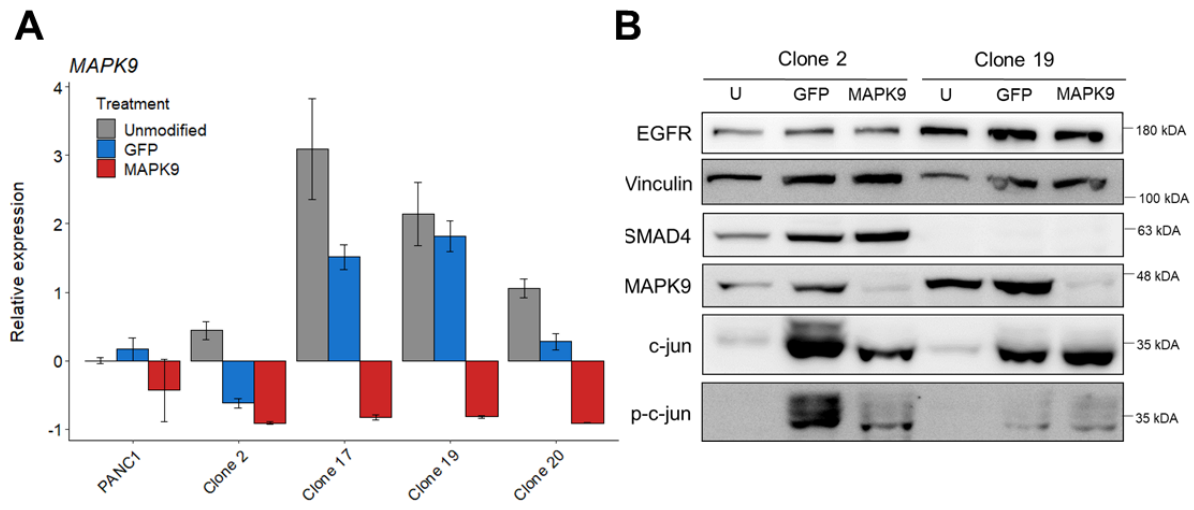


Figure 3.53: Overview of *MAPK9* mRNA and protein levels after siRNA knockdown. A) Relative *MAPK9* mRNA expression levels in unmodified cells, cells treated with the negative control GFP siRNA (GFP), and cells incubated with MAPK9 siRNA (MAPK9). MAPK9 siRNA decreased *MAPK9* mRNA levels significantly in all clones and in the PANC1 cell line. ΔΔCT values normalised based on *GAPDH* CT and ΔCT values of unmodified PANC1. **B)** The western blot of clone 2 and clone 19 showed a significant decrease in MAPK9 after MAPK9 siRNA treatment as well, and it showed effect on the EGFR or SMAD4 protein levels. Activation of downstream effector c-jun through phosphorylation occurred in the negative control as well; therefore, it seemed to be an unspecific stress response to siRNA transfection rather than the result of a change in jun-kinase signalling.

No decrease in erlotinib resistance was observed in *SMAD4*- clone 19 following MAPK9 knockdown (Figure 3.54).

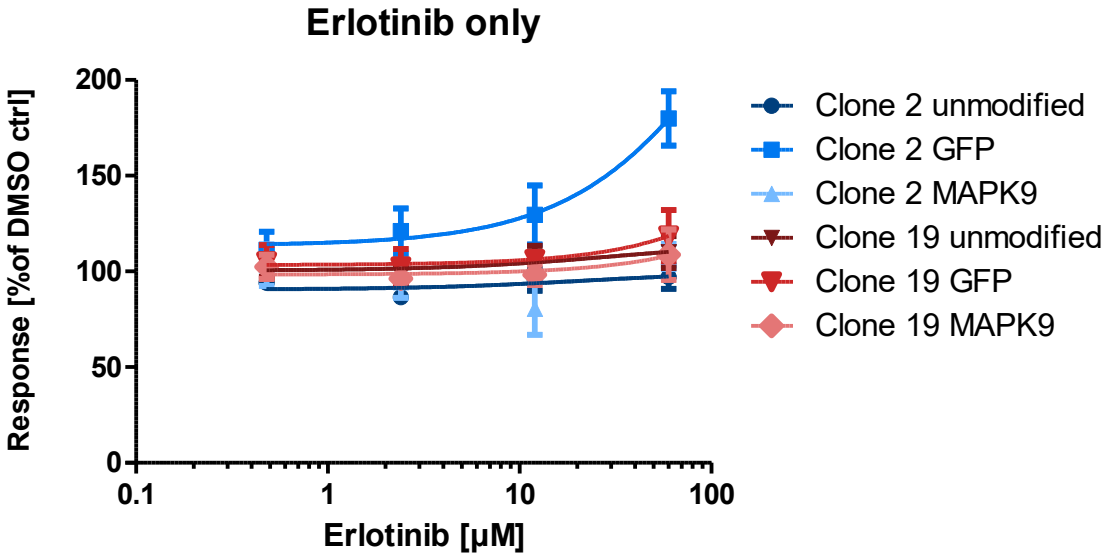


Figure 3.54: Chemosensitivity in *SMAD4*^{alt} *MAPK9*^{low} cells. For each clone and treatment, a dose-response curve is fitted through normalized WST absorption values at different erlotinib concentrations. No difference in cell viability was noted in *MAPK9* knockdown cells.

4 Discussion

This thesis provides a detailed description of the genetic heterogeneity of R0-resected PDAC patients. The homogenous nature of both tumour stage and treatment, the detailed clinical annotations, the large cohort size, and the long follow-up presented a unique opportunity to identify subtle molecular biological patterns previously missed. While patient-specific genotyping has become more feasible over the past decade, targeted therapy options remain limited in PDAC. The most established subgrouping approaches for PDAC are based on the clustering of whole transcriptome sequencing data[102,107,108,181,182]. While these clusters have offered points of action for targeted cancer treatment, the downside of this approach is that high-quality patient material with high tumour content is needed for whole transcriptome sequencing, which is problematic in clinical praxis. All analyses in this thesis are based on targeted sequencing approaches of FFPE samples with a wide range of tumour content. The results are therefore more easily and cost effectively reproducible, which improves their clinical applicability.

4.1 Targeted molecular biological characterisation of FFPE tumour samples

In translational cancer research of solid tumours, the need for patient tissue that can be easily stored and transported led to the wide-spread use of FFPE for tissue preservation. The main disadvantage of this method is strong fragmentation of nucleic acids. For the CO-2016 cohort, I implemented a strict quality control regime to assess the amount of DNA and RNA fragmentation and adjusted the subsequent methodology accordingly. For library preparation, DNA samples with high fragmentation rates received a less intense ultra-sound fragmentation regime and additional PCR cycles to ensure sufficient amplicons in the required size. For RNA expression analysis, I chose the multiplexed single molecule hybridisation approach of the nCounter Technology. While the targeted nature of this method limits the genes that can be analysed to about 700 pre-selected ones, it is significantly less dependent on RNA fragmentation rates than RNA sequencing approaches. Similarly, in using MLPA for CNA detection, I limited the number of analysed genes to account for the low quality of some samples. To ensure that the final variants were true somatic mutations, I validated a high number of both SNVs and CNAs, reaching a final validation rate of muss >95 %.

Another major hurdle of most sequencing studies on PDAC, including our own, is the low tumour cell content within resected samples. While nucleic acids were extracted from areas rich in tumour cells using macro dissection, the final cell bulk still contained tumour stroma cells. With no matched normal control samples available, I used a cohort of 20 independent FFPE tissue samples and strict filtering criteria to exclude SNPs and sequencing errors. I also manually checked variants, particularly those with VAFs that significantly deviated from the

overall VAFs in each patient, to exclude germline variants. In both approaches, the targeted nature of my NGS approach was helpful as well. I could exclude all variants not found within the genes I included in the panel, thereby concentrating on PDAC and cancer-driver genes in general, which were more likely to contain reliable genetic variations with functional consequences.

4.2 In-depth analysis of PDAC based on multiple data layers increase insight into disease biology

Over the last decade several large sequencing studies have widely increased our knowledge about the genetic background underlying PDAC tumorigenesis or cancerogenesis. Genetic aberrations, including mutations and CNAs in four core genes, *KRAS*, *TP53, CDKN2A* and *SMAD4* have been repeatedly described. They are complemented by a long tail of genes affected in less than 10% of PDAC patients[101,102,183]. A similar alteration pattern was seen in the CO-2016 cohort. Mutation type and frequencies are similar to results from Bailey et al.[102] and PDAC-typical CNA frequencies such as focal deletions in 9p21 (*CDKN2A*), amplifications of 8q24 (*MYC*) and deletions of the chromosome 6 were detected[184]. This was a good indication that the quality of my sequencing data and the filter criteria I applied were adequate for reliable variant detection. Additionally, the genes I chose to include in the custom PDAC panel could be used to recreate known genetic patterns in PDAC, giving me a solid data base for subsequent analyses.

Recently, more in-depth analyses have revealed additional complex interactions. For instance, in addition to being mutated in over 90% of PDAC patients, *KRAS* has also repeatedly been found to be amplified. Amplification of the mutated allele results in increased tumour promotion potential[185]. A similar trend was observed in the CO-2016 cohort; patients with a *KRAS* G12D mutation and an amplified *KRAS* gene exhibited an increased percentage of distant metastasis compared to G12D mutated patients without amplification (71.5 % vs 48 %), though this failed to reach statistical significance. PDAC is a highly heterogeneous disease and several current classification systems exist to subgroup PDAC patients based on their expression signatures[102,107,108]. These partly overlapping signatures have shown to be associated with distinct clinical outcomes and differing chemotherapy responses[186]. Furthermore, earlier this year, Chan-Seng-Yue et al. showed that the constellation of genomic aberrations in the tumour gives rise to the molecular subtype, and that disease heterogeneity is due to ongoing genomic instability during progression. In line with this, they were able to further subdivide the two most accepted subtypes, Basal-like and Classical, and match them to either resectable or advanced disease[187]. An even more detailed picture of the heterogeneity underlining PDAC and its precursors was produced by the first single-cell sequencing approaches. As a result, a refined malignant progression model was proposed with a tremendous shift in micro environmental

cell populations over time[188]. As many as ten main cell clusters with widely different expression signatures and malignancies were identified within the tumour tissue. Only one appeared to be directly involved in cancer progression[189]. Additional multi-omics studies have both increased our understanding of PDAC and shown a multitude of complex interactions that we do not yet completely understand. Epigenetics signatures, associated with the classical and basal subtype, have helped identify key transcription factors. When modified, these transcription factors might facilitate subtype switches in PDAC cells[190]. Several long-noncoding RNAs, which are known to interact with epigenetic and posttranslational modifications, have been associated with distinct clinical prognosis and can be found early in bodily fluids, making them a possible diagnostic target[191]. Finally, analyses of the PDAC proteome provided additional understanding of the complex nature of the PDAC microenvironment and proposed several potential prognostic biomarkers[192]. To account for the interplay between these complex networks, a trend toward a combined approach was observed. Our targeted experimental set-up prevented the classification into known expression signatures and the degradation of FFPE samples, made both epigenetic and proteome analysis impossible. Nevertheless, we did use three data layers (SNV, CNA, mRNA expression) to identify complex biological structures within 283 patients. Non-negative matrix factorization clustering based on the SNV and CNA of the patients revealed four distinct genetic signatures, which in turn could be used to subgroup patients via hierarchical clustering. The resulting clusters not only helped to identify a possible starting point in the search for predictive biomarkers for erlotinib, but they identified two very different genetic backgrounds driving PDAC development. Patients in cluster 5 had only few alterations in addition to the initiating *KRAS* mutations. They had a favourable prognosis and increased MAPK pathway expression as well as high levels of *PTEN*. This implies that these patients' tumours are driven by the early PDAC events of hyperactive *KRAS*, which infers a better prognosis. Contrast cluster 2 showed an enrichment of deletions in three major tumour suppressors (*PTEN*, *RB1*, *BRCA2*). This led to the uncontrolled activation of the PI3K-Akt signalling pathway, shown by the increased expression of its receptors and effectors. Therefore, tumours in cluster 2 did not appear to rely solely on the initial MAPK pathway deregulation but have evolved to deregulate a second cell proliferation pathway – with the effect of decreased DFS and OS. While these findings cannot be used directly in clinical practice, they open up another starting point for research into the biological processes underlying PDAC progression. Results from the NMF clustering with more direct clinical implications are potential novel points of action. Especially cluster 2 patients, who exhibited the prognosis of all patients, showed several potentially actionable alterations. For example, the oral pan-AKT Inhibitor MK-2206 has shown to decrease tumor size and CA19-9 levels in PDAC patients with *PTEN* loss[193], and the PARP inhibitor olaparib prolonged progression-free

survival in metastatic PDAC patients with *BRCA1/2* mutations[194]. Incidences for both alterations were significantly increased in cluster 2 patients.

4.3 Predicting clinical outcome of resectable PDAC

Predicting the clinical course of disease is an important step in cancer diagnostics. Patients with a high risk of recurrence, early dissemination, or fast tumour progression who require aggressive treatment regimens must be distinguished from low-risk patients for whom disease monitoring and treatment of symptoms might be enough. This is necessary for maximising health care effectiveness and minimising patient suffering. Historically, pathological classification systems like tumour size and differentiation status have acted as the basis of this estimation. In recent decades, molecular biomarkers have been used increasingly to refine outcome prediction. These prognostic biomarkers can be single genetic factor alterations like *BRAF* V600E mutations in advanced microsatellite stable colorectal cancer[195]. They can also be oestrogen/progesterone receptor statuses in breast cancer[196] or a multi gene panel or Oncotype DX for breast cancer[197]. The search for prognostic biomarkers is a major goal in all cancer entities, but even for promising candidates, the necessity for expensive clinical trials can make implementation challenging. The only FDA-approved biomarker for pancreatic cancer is CA19-9, which can be used to validate PDAC diagnosis in symptomatic patients. While the serum biomarker can provide important prognostic information and non-specific expression in several benign and malignant diseases, false negative results for some genotypes and an increase in false positive results in the presence of obstructive jaundice severely limit its universal applicability[198]. The only other gene-based prediction tool used in clinical practice are mutation panels that include genes specific for hereditary PDAC (e.g. *BRCA1/2, PALB2, ATM)*. They can be used to assess PDAC risk in patients with a family background of the disease[150]. Several other prognostic biomarkers have been implicated in PDAC but none have been validated in clinical trials thus far. I did encounter several genetic alterations with a negative effect on OS in the CO-2016 cohort. They included *KRAS* Q61 (n=15, log rank p=0.05), *PIK3CG* (n=12, log rank p=0.02), and *TP53* missense (n=68, log rank p=0.04) mutations. While *KRAS* mutations in general are often associated with decreased OS[111], *KRAS* Q61 mutations have previously been found in patients with a better prognosis[183]. *TP53* mutations have also been discussed as negative prognostic markers in PDAC, with differing results. However, no specific effect of missense mutations was described[113]. While there are clear indications that *PIK3CG* plays an important role in PDAC progression[199], it has not yet been implicated as a potential prognostic biomarker. Due to these controversial reports and the fact that none of the associations were found to be significant after adjustment for multiple testing, their validity as potential prognostic biomarkers remains doubtful. Rather, I focused on expression levels of genes found to be significantly associated with OS and

combined them with established clinical factors to create a genetic risk score for clinical outcome predictions in PDAC. There have been previous attempts to combine clinical parameters and create a nomogram for resectable[200], locally advanced[201], and metastatic PDAC[202]. While these nomograms more accurately predict prognosis than single factors do, they have not yet been widely used in clinical practice, and international guidelines do not recommend the use of nomograms as predictive tools to guide treatment decisions[203]. Therefore, further refinement through inclusion of molecular markers is necessary.

When only three expression markers, all of which had previously been associated with poor prognosis in PDAC[170], were incorporated, it allowed for the identification of three distinct risk groups with median survival differences of up to three years. Of note, high-risk group patients showed extremely poor outcomes with median DFS and OS of 7 and 13 months respectively. Similar survival outcomes have been reported for metastatic PDAC patients treated with FOLFIRINOX (DFS: 6 and OS: 11 months)[90]. The three survival groups identified by the newly proposed risk score were compared against survival groups assigned using a simple clinical score composed of N-stage and grading. In so doing, I observed an important difference in patient assignment: between 17 to 28 patients shifted groups between both scores (e.g., 17 patients with an intermediate risk according to the clinical score were defined as high-risk) when including gene expression counts, leading to a refined prediction for the course of disease. Incorporation of molecular parameters was especially important to clearly distinguish the high-risk patients. Interestingly, the three survival groups also showed significant differences in the number of patients with relapse, relapse type (local vs distant), organ distribution (liver vs lung), and DFS. Even though all patients were R0-resected and received adjuvant chemotherapy, patients in the low-risk group relapsed significantly less often, and, if they did, they had a more local relapse pattern. No apparent genetic difference between the subgroups could be found. While high-risk patients had significantly more alterations (mean alterations: 23.35 ± 10.74) than low-risk patients (14.23 ± 13.04; Unpaired *t* test p = 0.0001), no single alteration was specifically enriched in either group. It would be interesting to further analyse whether the differences in survival and relapse are based on variances in reaction to the adjuvant chemotherapy or underlying biological differences in the tumour itself which lead to a faster relapse and a more widespread metastasis pattern. A potentially gemcitabine-resistant subgroup might need a different adjuvant treatment or should not get adjuvant chemotherapy at all to avoid side effects.

While the fact that the risk score could be replicated in a second independent cohort is promising, further validation for instances in the form of a basket study is necessary before the score can be used for treatment stratification.

4.4 Interactions underlying *SMAD4* - *MAPK9* dependent erlotinib sensitivity in PDAC patients

In contrast to other solid tumours, PDAC has proven to be especially resistant to any targeted treatment strategy. In 2007, Moore et al. showed a significant increase in OS for advanced metastatic PDAC patients after adding erlotinib to the standard treatment regime[94]. Even that small increase in OS was welcome. Erlotinib is a small molecule inhibitor of EGFR, which competitively and reversely binds and blocks the ATP binding site of the catalytic domain. Subsequently, its autophosphorylation is prevented and downstream signalling pathways are weakened[204]. Differing from other anti-EGFR therapeutics like gefitinib, erlotinib has shown to also strongly bind additional protein kinases[205]. Erlotinib is FDA-approved for the treatment of patients with advanced, chemotherapy-resistant, non-small-cell lung carcinoma (NSCLC) in combination with gemcitabine for advanced pancreatic cancer[206]. In contrast to NSCLC wherein EGFR mutations and deletions were shown to increase erlotinib sensitivity, no predictive biomarker for erlotinib treatment is known in PDAC to date. Additionally, the small, two-week increase in advanced PDAC patients makes its use in clinical practice scarce. The CONKO-005 study was a randomized multicentre phase III study that tested the addition of erlotinib to gemcitabine as an adjuvant treatment regime for R0-resected PDAC patients to see whether the effect shown by Moore et al[94] could be replicated in local disease. Though no significant difference in DFS or OS was found between the two treatment arms, a late split between both curves indicated the presence of a small PDAC subgroup that might benefit from erlotinib treatment[142].

In this thesis, targeted genotyping of 293 patients from the CONKO-005 study revealed SMAD4 to be a key component in regulating erlotinib sensitivity in PDAC. In contrast to the broad scientific consensus of *SMAD4* levels as negative prognostic biomarker in PDAC[207], *SMAD4*^{alt} patients did not show a significantly shorter DFS or OS in the CO-2016 cohort. This could be explained by the fact that the negative effect of *SMAD4* alteration in this cohort was masked by the difference in treatment. When accounting for this by either a multivariate cox regression analysis or comparing *SMAD4*^{alt} and WT patients in the gemcitabine arm alone, *SMAD4* alterations did show a significant impact on OS and DFS. In contrast, within the erlotinib treatment arm, the effect of *SMAD4* alterations was negated leading to a survival curve of *SMAD4*^{alt} to be comparable to that of WT patients without erlotinib treatment. Interestingly, in two previous studies, where PDAC patients were treated with EGFR inhibitors, low SMAD4 immunostaining was associated with a metastatic relapse but not with a decrease in OS[208,209].

The addition of *MAPK9* expression levels, a molecule found to be significantly less expressed in *SMAD*^{alt} patients, further specified patients who benefitted from erlotinib treatment. There is a complex signalling network connecting TGFβ and JNK signalling in PDAC. TGFβ is a

multifaceted signalling molecule overexpressed in several cancers[210]. After binding to its receptor, TGFBR1 and 2 form heterotetramers and relay the signal either via a canonical SMAD dependent pathway or by interacting with different major signalling pathways in a non-canonical signal transduction. The dual nature of TGFβ signalling in PDAC tumour initiation and progression has been the focus of many analyses. It has led to the model that, while the canonical pathway act as tumour suppressors in early stages, TGFβ becomes a main driver of EMT and metastasis formation in advanced disease[211]. SMAD4 has shown to be the most prominent regulator of this functional switch. In healthy cells, TGFβ maintains homeostasis through immune modulation and blockage of mitogenic growth signals through SMAD4[212,213]. Loss of SMAD4 via point mutation or deletion of the 18q21 chromosomal segment reduces the TGFβ/SMAD4-dependent cell cycle arrest and apoptosis in pancreatic cells[214,215] and conversely increases tumour-–progressive TGFβ signalling through SMAD4 independent pathways. In advanced cancer, overexpression of TGFβ activates Ras/Erk, PI3K/Akt, p38 MAPK, and Rho-GTPase pathways, which all play a role in tumorigenesis[216]. These pathway interconnections lead to SMAD4 depending changes to anti-EGFR treatments that could be shown in different tumour entities. Treatment with the anti-EGFR antibody cetuximab increases expression of TGFβ in human head and neck squamous cell carcinoma (HNSCC) cells, and inhibition with a TGFβ-blocking antibody increased the anti-tumour efficacy in mice-bearing xenografts[217]. In a panel of human cancer cells, Manole et al. showed that acquired resistance to erlotinib is modulated by the activation of the JNK pathway[177], and that deletion of SMAD4 led to cetuximab resistance regulated by increased pJNK and pMAPK levels in HNSCC cells[218].

To validate the findings from the PDAC patient cohort and to be able to study the underlying mechanisms, I modelled the *SMAD4*^{alt} *MAPK9*^{low} phenotype *in vitro* using *SMAD4* naïve PANC1 cells. Through this process, I was able to knockdown functional protein production of SMAD4 completely via CRISPR/Cas9 genetic modification and to significantly reduce *MAPK9* expression via siRNA. However, no subsequent increase in erlotinib-based cell toxicity could be seen. A similar effect was found by Chen et al., where stable *SMAD4* knockout in PANC1 cells did not increase sensitivity to gefitinib treatment[219]. Interestingly, while almost all *SMAD4* negative cell lines showed a strong response to erlotinib treatment, both PANC1 and BXPC3 did not show any erlotinib-based cell toxicity. It is of note that the gemcitabine sensitivity significantly increased in *SMAD4* KO cells. This effect is controversial in literature, where both an increase[220] and a decrease[219,221] of gemcitabine sensitivity in *SMAD4* negative cells has been described. Since an increase in erlotinib-dependent cytotoxicity in *SMAD4* KO cells was not the basis for the prolonged DFS and OS in the CONKO-005 cohort, a different effect must have been responsible. *SMAD4* genetic status has been described repeatedly as a marker for distant, metastasis-driven relapse[222,223]. One molecular basis for metastasis formation is the

EMT. The EMT can be a normal physiological process necessary for embryonic development where epithelial cells lose their cell polarity and cell-to-cell contact to acquire motility and the ability to invade foreign sides. In cancer, this process is deregulated and allows the dissemination of the tumour cells throughout the body. TGFβ is one of the key regulators of this transition process[224]. In *SMAD4* WT cells, the SMAD3/4 complex induces the transcription of *SNAIL* and *ZEB* transcription factors, both activators of EMT, as well as downregulates the expression of E-cadherin[216]. In cells with impaired SMAD4, TGFβ signalling shifts to non-canonical signal transduction, leading to the activation of the JNK/p38 MAPK pathway which can in turn induce EMT[225]. Recently, the SMAD4-dependent EMT has been linked to parallel induction of apoptosis, resulting in a lethal EMT phenotype which might explain the dual nature of TFGβ signalling[215]. Some hints in the clinical data of the CO-2016 cohort point to changes in metastasis formation during erlotinib treatment as the critical difference between *SMAD4*^{alt} *MAPK9*^{low} patients and the rest. In *SMAD4*^{alt} *MAPK9*^{low} patients treated with erlotinib, a trend suggesting lower relapse occurrence and a shift of distant metastasis toward lung only – a phenotype previously associated with improved outcome[171] – was observed. To test this in my *in vitro* model, I did a wound scratch assay of *SMAD4* KO and WT clones treated with or without erlotinib. *SMAD4* KO cells had a decrease in cell motility and erlotinib treatment further decreased the effect. Some decrease of cell motility could also be seen in *SMAD4* WT but to a lesser extent. The impact of SMAD4 in TGFβ-induced EMT and cell migration has been shown previously. Chen et al. detected an increase in expression of EMT markers and pMAPK in *SMAD4* KO PANC1 cells[219]. SMAD4 inhibition led to decreased cell migration via suppression of JNK activity in human PDAC cell lines[226]. Loss of SMAD4 was associated with increased invasion even in human cancer organoids [227]. The underlying signalling network is complex and only partly understood. Two molecules with a potential role in this interaction are KIT and MECOM. Both were expressed significantly differentially in *SMAD4*^{alt} patients and were overexpressed in all three *SMAD4* KO clones. The proto-oncogene receptor tyrosine kinase KIT is an activator of the MAPK pathway and therefore a potential substitute for EGFR. Its role in PDAC has not been fully analysed, but high c-KIT expression levels have been associated with an improved outcome in resectable PDAC[228]. MECOM is a tumour suppressor that regulates both the TGFβ signalling pathway by binding and blocking SMAD3 and the MAPK pathway by binding and blocking jun-kinases. In the human protein atlas, it is listed as a strong negative prognostic marker and was shown to promote the KRAS pathway trough expression of one of its main inhibitors in early PDAC carcinogenesis[170,229]. A more detailed analysis of the interaction between erlotinib, the TGFβ, and the JNK signalling pathway within the *in vitro* model are necessary to clarify the role MECOM and KIT play in SMAD4-driven EMT. Currently, we know that *SMAD4*^{alt} *MAPK9*^{low} patients had a significantly reduced DFS and OS when erlotinib was part of their adjuvant chemotherapy. Modelling the genetic set-up

for *in vitro* has shown no evidence for increased cytotoxicity of that erlotinib is the cause of this effect. Differences in metastasis patterns and previous findings from literature point to changes in relapse pattern as one possible explanation.

Another important pillar in the carcinogenesis of PDAC is the tumour-immune microenvironment. Li et al.[230] showed that potent epigenetic regulator lysine demethylase 3A (KDM3A) interacts with Krueppel-like factor 5 (KLF5) and SMAD4 to regulate the expression of EGFR. In their mouse model, both *SMAD4* deletion and EGFR inhibition led to an immunogenic, active, and T-cell-enriched phenotype. This would suggest a synergy between erlotinib and the *SMAD4*^{alt} phenotype, creating a T cell-rich microenvironment that might have some anti-tumour efficacy. The DE analysis of patients from the CO-2016 cohort showed several players of the adaptive and innate immune response to be significantly over- (*RAC1, NFkB1, DTX4, CREBBP*) or under-expressed (*COL1A2, UBB, PLD1, IKBKB, RAF1, MAPK9*) in *SMAD4*^{alt} patients. This could suggest a difference in the tumour-immune microenvironment between both subgroups. Nevertheless, the effect of SMAD4 on tumour immunity is not yet fully understood. Bali et al.[231] observed a non-significant trend toward a cytolytic low (T cell non-inflamed) phenotype associated with copy number loss or mutation in SMAD4, contradicting Li's findings.

The complex nature of the PDAC microenvironment, with its dense stroma and strong immune suppression, would require a more complex 3D model to validate the use of *SMAD4* alterations and *MAPK9* expression level as predictive biomarker for erlotinib sensitivity..

5 Conclusion and Outlook

The molecular biological characterisation of 293 patients previously enrolled in the CONKO-005 study resulted in a comprehensive and reliable data set with three layers of data (SNV, CNA, expression) combined with detailed clinical information for each patient. Adequate quality control measures, experimental adjustments, and a stringent validation regime were employed to guarantee excellent data quality despite highly fragmented FFPE material. As a first comprehensive analysis, combination of the molecular biological data layers via NMF and subsequent DE analysis identified two PDAC patient subgroups with distinct biological characteristics and significant differences in clinical outcomes. While they could not be directly transferred into a treatment strategy, they offered insight into the biological background of PDAC survival groups and contained several targetable alterations. My second finding, the genetic risk score based on two clinical and three expression markers, is more directly applicable in clinical practice. With its help, patients can be sorted into three survival groups with significant differences in DFS, OS, and the metastasis pattern. I was able to reproduce this strong split among patient risk groups in a second independent cohort; nevertheless, both analyses have been retrospective and need further validation in a prospect study. The third and final major finding of this thesis was a significant increase in DFS and OS in *SMAD4*^{alt} *MAPK9*^{low} patients who were treated with erlotinib. It offers a potential predictive biomarker combination for the adjuvant treatment of resectable PDAC. I could show that the changes in erlotinib sensitivity are not based on an increased cell toxicity of erlotinib treatment in *SMAD4* KO *MAPK9*^{low} PDAC cells. While the patient cohort and in the *in vitro* model did at times indicate that changes to EMT and subsequent relapse patterns might be the basis instead, the underlying signalling networks are still largely unidentified. Especially in PDAC, the dense tumour stroma and widespread immune suppression are responsible for many effects in patients' courses of disease and treatment outcomes. TGFβ signalling is a key regulator of the tumour environment in PDAC. Therefore, several potential mechanisms underlying the effect of erlotinib treatment seen in the patient cohort can only be studied in a more complex 3D cell culture or a mouse model. Nevertheless, this thesis offers another piece of the puzzle that is the complex biology of PDAC and opens up several new points of action for the treatment of one of the most deadly tumours in the world.

References

1. Hoyer K, Hablesreiter R, Inoue Y, et al: A genetically defined signature of responsiveness to erlotinib in early-stage pancreatic cancer patients: results from the CONKO-005 trial. EBioMedicine, accepted for publication: 19.03.2021
2. Harris RE: Epidemiology of Chronic Disease: Global Perspectives, Jones & Bartlett Learning, 2013
3. Ryan DP, Hong TS, Bardeesy N: Pancreatic Adenocarcinoma. New England Journal of Medicine 371:1039-1049, 2014
4. Neoptolemos JP, Abbruzzese J, Urrutia RA, et al: Pancreatic Cancer, Springer, 2010
5. Feldman M, Friedman LS, Brandt LJ: Sleisenger and Fordtran's Gastrointestinal and Liver Disease E-Book: Pathophysiology, Diagnosis, Management, Elsevier Health Sciences, 2015
6. Al-Hader A, Al-Rohil RN, Han H, et al: Pancreatic acinar cell carcinoma: A review on molecular profiling of patient tumors. World journal of gastroenterology 23:7945-7951, 2017
7. logika600: https://creativemarket.com/nmfotograf/1498100-Pancreas-and-duodenum-location, 28-10-2020
8. Öberg K, Knigge U, Kwekkeboom D, et al: Neuroendocrine gastro-entero-pancreatic tumors: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. Annals of Oncology 23:vii124-vii130, 2012
9. Burns WR, Edil BH: Neuroendocrine Pancreatic Tumors: Guidelines for Management and Update. Current Treatment Options in Oncology 13:24-34, 2012
10. Modolell I, Guarner L, Malagelada JR: Vagaries of clinical presentation of pancreatic and biliary tract cancer. Ann Oncol 10 Suppl 4:82-4, 1999
11. Porta M, Fabregat X, Malats N, et al: Exocrine pancreatic cancer: symptoms at presentation and their relation to tumour site and stage. Clin Transl Oncol 7:189-97, 2005
12. Chari ST, Leibson CL, Rabe KG, et al: Probability of pancreatic cancer following diabetes: a population-based study. Gastroenterology 129:504-11, 2005
13. Edge SB, Compton CC: The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. Ann Surg Oncol 17:1471-4, 2010
14. Jalanko H, Kuusela P, Roberts P, et al: Comparison of a new tumour marker, CA 19-9, with alpha-fetoprotein and carcinoembryonic antigen in patients with upper gastrointestinal diseases. Journal of Clinical Pathology 37:218-222, 1984
15. Dong Q, Yang XH, Zhang Y, et al: Elevated serum CA19-9 level is a promising predictor for poor prognosis in patients with resectable pancreatic ductal adenocarcinoma: a pilot study. World J Surg Oncol 12:171, 2014
16. Marchegiani G, Andrianello S, Malleo G, et al: Does Size Matter in Pancreatic Cancer?: Reappraisal of Tumour Dimension as a Predictor of Outcome Beyond the TNM. Ann Surg 266:142-148, 2017
17. Goh SK, Gold G, Christophi C, et al: Serum carbohydrate antigen 19-9 in pancreatic adenocarcinoma: a mini review for surgeons. ANZ J Surg 87:987-992, 2017
18. Azizian A, Rühlmann F, Krause T, et al: CA19-9 for detecting recurrence of pancreatic cancer. Scientific Reports 10:1332, 2020
19. Cong L, Liu Q, Zhang R, et al: Tumor size classification of the 8th edition of TNM staging system is superior to that of the 7th edition in predicting the survival outcome of pancreatic cancer patients after radical resection and adjuvant chemotherapy. Scientific Reports 8:10383, 2018
20. Bray F, Ferlay J, Soerjomataram I, et al: Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians 68:394-424, 2018
21. Rawla P, Sunkara T, Gaduputi V: Epidemiology of Pancreatic Cancer: Global Trends, Etiology and Risk Factors. World journal of oncology 10:10-27, 2019

22. Wong MCS, Jiang JY, Liang M, et al: Global temporal patterns of pancreatic cancer and association with socioeconomic development. Sci Rep 7:3165, 2017

23. Malvezzi M, Carioli G, Bertuccio P, et al: European cancer mortality predictions for the year 2016 with focus on leukaemias. Ann Oncol 27:725-31, 2016

24. Duggan MA, Anderson WF, Altekruse S, et al: The Surveillance, Epidemiology, and End Results (SEER) Program and Pathology: Toward Strengthening the Critical Relationship. The American journal of surgical pathology 40:e94-e102, 2016

25. McGuigan A, Kelly P, Turkington RC, et al: Pancreatic cancer: A review of clinical diagnosis, epidemiology, treatment and outcomes. World J Gastroenterol 24:4846-4861, 2018

26. Lee D-H, Jang J-Y, Kang JS, et al: Recent treatment patterns and survival outcomes in pancreatic cancer according to clinical stage based on single-center large-cohort data. Annals of hepato-biliary-pancreatic surgery 22:386-396, 2018

27. Siegel RL, Miller KD, Jemal A: Cancer statistics, 2017. CA: A Cancer Journal for Clinicians 67:7-30, 2017

28. Rahib L, Smith BD, Aizenberg R, et al: Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. Cancer Res 74:2913-21, 2014

29. Pourshams A, Sepanlou SG, Ikuta KS, et al: The global, regional, and national burden of pancreatic cancer and its attributable risk factors in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. The Lancet Gastroenterology & Hepatology 4:934-947, 2019

30. Kuzmickiene I, Everatt R, Virviciute D, et al: Smoking and other risk factors for pancreatic cancer: a cohort study in men in Lithuania. Cancer Epidemiol 37:133-9, 2013

31. Pelucchi C, Galeone C, Polesel J, et al: Smoking and body mass index and survival in pancreatic cancer patients. Pancreas 43:47-52, 2014

32. Mizuno S, Nakai Y, Isayama H, et al: Smoking, family history of cancer, and diabetes mellitus are associated with the age of onset of pancreatic cancer in Japanese patients. Pancreas 43:1014-7, 2014

33. Wang YT, Gou YW, Jin WW, et al: Association between alcohol intake and the risk of pancreatic cancer: a dose-response meta-analysis of cohort studies. BMC Cancer 16:212, 2016

34. Samokhvalov AV, Rehm J, Roerecke M: Alcohol Consumption as a Risk Factor for Acute and Chronic Pancreatitis: A Systematic Review and a Series of Meta-analyses. EBioMedicine 2:1996-2002, 2015

35. Raimondi S, Lowenfels AB, Morselli-Labate AM, et al: Pancreatic cancer in chronic pancreatitis; aetiology, incidence, and early detection. Best Pract Res Clin Gastroenterol 24:349-58, 2010

36. Aune D, Greenwood DC, Chan DS, et al: Body mass index, abdominal fatness and pancreatic cancer risk: a systematic review and non-linear dose-response meta-analysis of prospective studies. Ann Oncol 23:843-52, 2012

37. Li D, Morris JS, Liu J, et al: Body mass index and risk, age of onset, and survival in patients with pancreatic cancer. Jama 301:2553-62, 2009

38. Schwartz GG, Reis IM: Is cadmium a cause of human pancreatic cancer? Cancer Epidemiol Biomarkers Prev 9:139-45, 2000

39. Kriegel AM, Soliman AS, Zhang Q, et al: Serum cadmium levels in pancreatic cancer patients from the East Nile Delta region of Egypt. Environ Health Perspect 114:113-9, 2006

40. Ojajärvi IA, Partanen TJ, Ahlbom A, et al: Occupational exposures and pancreatic cancer: a meta-analysis. Occup Environ Med 57:316-24, 2000

41. Midha S, Chawla S, Garg PK: Modifiable and non-modifiable risk factors for pancreatic cancer: A review. Cancer Lett 381:269-77, 2016

42. Arnold LD, Patel AV, Yan Y, et al: Are racial disparities in pancreatic cancer explained by smoking and overweight/obesity? Cancer Epidemiol Biomarkers Prev 18:2397-405, 2009

43. Pernick NL, Sarkar FH, Philip PA, et al: Clinicopathologic analysis of pancreatic adenocarcinoma in African Americans and Caucasians. Pancreas 26:28-32, 2003

44. Stevens RJ, Roddam AW, Beral V: Pancreatic cancer in type 1 and young-onset diabetes: systematic review and meta-analysis. Br J Cancer 96:507-9, 2007

45. Huxley R, Ansary-Moghaddam A, Berrington de González A, et al: Type-II diabetes and pancreatic cancer: a meta-analysis of 36 studies. Br J Cancer 92:2076-83, 2005

46. Hruban RH, Petersen GM, Goggins M, et al: Familial pancreatic cancer. Ann Oncol 10 Suppl 4:69-73, 1999

47. Vincent A, Herman J, Schulick R, et al: Pancreatic cancer. Lancet 378:607-20, 2011

48. Wolpin BM, Chan AT, Hartge P, et al: ABO blood group and the risk of pancreatic cancer. J Natl Cancer Inst 101:424-31, 2009

49. Basturk O, Hong SM, Wood LD, et al: A Revised Classification System and Recommendations From the Baltimore Consensus Meeting for Neoplastic Precursor Lesions in the Pancreas. Am J Surg Pathol 39:1730-41, 2015

50. Guerra C, Collado M, Navas C, et al: Pancreatitis-induced inflammation contributes to pancreatic cancer by inhibiting oncogene-induced senescence. Cancer cell 19:728-739, 2011

51. Collado M, Gil J, Efeyan A, et al: Senescence in premalignant tumours. Nature 436:642-642, 2005

52. Matsuda Y, Furukawa T, Yachida S, et al: The Prevalence and Clinicopathological Characteristics of High-Grade Pancreatic Intraepithelial Neoplasia: Autopsy Study Evaluating the Entire Pancreatic Parenchyma. Pancreas 46:658-664, 2017

53. Notta F, Hahn SA, Real FX: A genetic roadmap of pancreatic cancer: still evolving. Gut, 2017

54. Lüttges J, Galehdari H, Bröcker V, et al: Allelic loss is often the first hit in the biallelic inactivation of the p53 and DPC4 genes during pancreatic carcinogenesis. Am J Pathol 158:1677-83, 2001

55. Maitra A, Adsay NV, Argani P, et al: Multicomponent analysis of the pancreatic adenocarcinoma progression model using a pancreatic intraepithelial neoplasia tissue microarray. Mod Pathol 16:902-12, 2003

56. Hruban RH, Goggins M, Parsons J, et al: Progression model for pancreatic cancer. Clinical Cancer Research 6:2969-2972, 2000

57. Haeno H, Gonen M, Davis MB, et al: Computational modeling of pancreatic cancer reveals kinetics of metastasis suggesting optimum treatment strategies. Cell 148:362-75, 2012

58. Yachida S, Jones S, Bozic I, et al: Distant metastasis occurs late during the genetic evolution of pancreatic cancer. Nature 467:1114-1117, 2010

59. Makohon-Moore AP, Zhang M, Reiter JG, et al: Limited heterogeneity of known driver gene mutations among the metastases of individual patients with pancreatic cancer. Nature Genetics 49:358, 2017

60. Notta F, Chan-Seng-Yue M, Lemire M, et al: A renewed model of pancreatic cancer evolution based on genomic rearrangement patterns. Nature advance online publication, 2016

61. Schlesinger Y, Yosefov-Levi O, Kolodkin-Gal D, et al: Single-cell transcriptomes of pancreatic preinvasive lesions and cancer reveal acinar metaplastic cells' heterogeneity. Nature Communications 11:4516, 2020

62. Qadir MMF, Álvarez-Cubela S, Klein D, et al: Single-cell resolution analysis of the human pancreatic ductal progenitor cell niche. Proceedings of the National Academy of Sciences 117:10876-10887, 2020

63. Erkan M, Hausmann S, Michalski CW, et al: The role of stroma in pancreatic cancer: diagnostic and therapeutic implications. Nat Rev Gastroenterol Hepatol 9:454-67, 2012

64. Neesse A, Algül H, Tuveson DA, et al: Stromal biology and therapy in pancreatic cancer: a changing paradigm. Gut 64:1476-84, 2015

65. Ayala G, Tuxhorn JA, Wheeler TM, et al: Reactive stroma as a predictor of biochemical-free recurrence in prostate cancer. Clin Cancer Res 9:4792-801, 2003

66. Ren B, Cui M, Yang G, et al: Tumor microenvironment participates in metastasis of pancreatic cancer. Molecular Cancer 17:108, 2018

67. Erkan M, Kurtoglu M, Kleeff J: The role of hypoxia in pancreatic cancer: a potential therapeutic target? Expert Rev Gastroenterol Hepatol 10:301-16, 2016

68. Vonderheide RH, Bayne LJ: Inflammatory networks and immune surveillance of pancreatic carcinoma. Curr Opin Immunol 25:200-5, 2013

69. Pylayeva-Gupta Y, Lee KE, Hajdu CH, et al: Oncogenic Kras-induced GM-CSF production promotes the development of pancreatic neoplasia. Cancer Cell 21:836-47, 2012

70. Ryschich E, Nötzel T, Hinz U, et al: Control of T-cell-mediated immune response by HLA class I in human pancreatic carcinoma. Clin Cancer Res 11:498-504, 2005

71. Mantovani A, Sica A, Sozzani S, et al: The chemokine system in diverse forms of macrophage activation and polarization. Trends Immunol 25:677-86, 2004

72. Hao NB, Lü MH, Fan YH, et al: Macrophages in tumor microenvironments and the progression of tumors. Clin Dev Immunol 2012:948098, 2012

73. Lonardo E, Frias-Aldeguer J, Hermann PC, et al: Pancreatic stellate cells form a niche for cancer stem cells and promote their self-renewal and invasiveness. Cell Cycle 11:1282-90, 2012

74. Sainz B, Jr., Martín B, Tatari M, et al: ISG15 is a critical microenvironmental factor for pancreatic cancer stem cells. Cancer Res 74:7309-20, 2014

75. Sainz B, Jr., Alcala S, Garcia E, et al: Microenvironmental hCAP-18/LL-37 promotes pancreatic ductal adenocarcinoma by activating its cancer stem cell compartment. Gut 64:1921-35, 2015

76. Kurahara H, Takao S, Maemura K, et al: M2-polarized tumor-associated macrophage infiltration of regional lymph nodes is associated with nodal lymphangiogenesis and occult nodal involvement in pN0 pancreatic cancer. Pancreas 42:155-9, 2013

77. Kurahara H, Takao S, Shinchi H, et al: Significance of lymphangiogenesis in primary tumor and draining lymph nodes during lymphatic metastasis of pancreatic head cancer. J Surg Oncol 102:809-15, 2010

78. Liu CY, Xu JY, Shi XY, et al: M2-polarized tumor-associated macrophages promoted epithelial-mesenchymal transition in pancreatic cancer cells, partially through TLR4/IL-10 signaling pathway. Lab Invest 93:844-54, 2013

79. Tang D, Zhang J, Yuan Z, et al: PSC-derived Galectin-1 inducing epithelial-mesenchymal transition of pancreatic ductal adenocarcinoma cells by activating the NF-κB pathway. Oncotarget 8:86488-86502, 2017

80. Wu YS, Chung I, Wong WF, et al: Paracrine IL-6 signaling mediates the effects of pancreatic stellate cells on epithelial-mesenchymal transition via Stat3/Nrf2 pathway in pancreatic cancer cells. Biochimica et Biophysica Acta (BBA) - General Subjects 1861:296-306, 2017

81. Tien YW, Wu YM, Lin WC, et al: Pancreatic carcinoma cells stimulate proliferation and matrix synthesis of hepatic stellate cells. J Hepatol 51:307-14, 2009

82. Grünwald B, Harant V, Schaten S, et al: Pancreatic Premalignant Lesions Secrete Tissue Inhibitor of Metalloproteinases-1, Which Activates Hepatic Stellate Cells Via CD63 Signaling to Create a Premetastatic Niche in the Liver. Gastroenterology 151:1011-1024.e7, 2016

83. Khorana AA, Mangu PB, Katz MHG: Potentially Curable Pancreatic Cancer: American Society of Clinical Oncology Clinical Practice Guideline Update Summary. J Oncol Pract 13:388-391, 2017

84. Oettle H, Neuhaus P, Hochhaus A, et al: Adjuvant chemotherapy with gemcitabine and long-term outcomes among patients with resected pancreatic cancer: the CONKO-001 randomized trial. Jama 310:1473-81, 2013

85. Conroy T, Hammel P, Hebbar M, et al: FOLFIRINOX or Gemcitabine as Adjuvant Therapy for Pancreatic Cancer. New England Journal of Medicine 379:2395-2406, 2018

86. Abbassi R, Schmid RM: Evolving Treatment Paradigms for Pancreatic Cancer. Visceral Medicine 35:362-372, 2019

87. Versteijne E, Suker M, Groothuis K, et al: Preoperative Chemoradiotherapy Versus Immediate Surgery for Resectable and Borderline Resectable Pancreatic Cancer: Results of the Dutch Randomized Phase III PREOPANC Trial. J Clin Oncol:Jco1902274, 2020

88. O'Reilly EM, Ferrone C: Neoadjuvant or Adjuvant Therapy for Resectable or Borderline Resectable Pancreatic Cancer: Which Is Preferred? Journal of Clinical Oncology 38:1757-1759, 2020

89. 3rd HAB, Moore MJ, Andersen J, et al: Improvements in survival and clinical benefit with gemcitabine as first-line therapy for patients with advanced pancreas cancer: a randomized trial. Journal of Clinical Oncology 15:2403-2413, 1997

90. Conroy T, Desseigne F, Ychou M, et al: FOLFIRINOX versus Gemcitabine for Metastatic Pancreatic Cancer. New England Journal of Medicine 364:1817-1825, 2011

91. Von Hoff DD, Ervin T, Arena FP, et al: Increased Survival in Pancreatic Cancer with nab-Paclitaxel plus Gemcitabine. New England Journal of Medicine 369:1691-1703, 2013

92. Pishvaian MJ, Bender RJ, Halverson D, et al: Molecular Profiling of Patients with Pancreatic Cancer: Initial Results from the Know Your Tumor Initiative. Clin Cancer Res 24:5018-5027, 2018

93. Canon J, Rex K, Saiki AY, et al: The clinical KRAS(G12C) inhibitor AMG 510 drives anti-tumour immunity. Nature 575:217-223, 2019

94. Moore MJ, Goldstein D, Hamm J, et al: Erlotinib Plus Gemcitabine Compared With Gemcitabine Alone in Patients With Advanced Pancreatic Cancer: A Phase III Trial of the National Cancer Institute of Canada Clinical Trials Group. Journal of Clinical Oncology 25:1960-1966, 2007

95. Chiaravalli M, Reni M, O'Reilly EM: Pancreatic ductal adenocarcinoma: State-of-the-art 2017 and new therapeutic strategies. Cancer Treatment Reviews 60:32-43, 2017

96. Hingorani SR, Zheng L, Bullock AJ, et al: HALO 202: Randomized Phase II Study of PEGPH20 Plus Nab-Paclitaxel/Gemcitabine Versus Nab-Paclitaxel/Gemcitabine in Patients With Untreated, Metastatic Pancreatic Ductal Adenocarcinoma. Journal of Clinical Oncology 36:359-366, 2018

97. Yarchoan M, Hopkins A, Jaffee EM: Tumor Mutational Burden and Response Rate to PD-1 Inhibition. N Engl J Med 377:2500-2501, 2017

98. Nevala-Plagemann C, Hidalgo M, Garrido-Laguna I: From state-of-the-art treatments to novel therapies for advanced-stage pancreatic cancer. Nature Reviews Clinical Oncology 17:108-123, 2020

99. Beatty GL, O'Hara MH, Lacey SF, et al: Activity of Mesothelin-Specific Chimeric Antigen Receptor T Cells Against Pancreatic Carcinoma Metastases in a Phase 1 Trial. Gastroenterology 155:29-32, 2018

100. Biankin AV, Waddell N, Kassahn KS, et al: Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. Nature 491:399-405, 2012

101. Waddell N, Pajic M, Patch A-M, et al: Whole genomes redefine the mutational landscape of pancreatic cancer. Nature 518:495-501, 2015

102. Bailey P, Chang DK, Nones K, et al: Genomic analyses identify molecular subtypes of pancreatic cancer. Nature 531:47-52, 2016

103. Ying H, Dey P, Yao W, et al: Genetics and biology of pancreatic ductal adenocarcinoma. Genes & Development 30:355-385, 2016

104. Solomon S, Das S, Brand R, et al: Inherited pancreatic cancer syndromes. Cancer J 18:485-91, 2012

105. Shi C, Hruban RH, Klein AP: Familial pancreatic cancer. Arch Pathol Lab Med 133:365-74, 2009

106. Chen F, Roberts NJ, Klein AP: Inherited pancreatic cancer. Chin Clin Oncol 6:58, 2017

107. Collisson EA, Sadanandam A, Olson P, et al: Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. Nat Med 17:500-503, 2011

108. Moffitt RA, Marayati R, Flate EL, et al: Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. Nature Genetics 47:1168, 2015

109. Birnbaum DJ, Bertucci F, Finetti P, et al: Molecular classification as prognostic factor and guide for treatment decision of pancreatic cancer. Biochimica et Biophysica Acta (BBA) - Reviews on Cancer 1869:248-255, 2018

110. Duconseil P, Gilabert M, Gayet O, et al: Transcriptomic Analysis Predicts Survival and Sensitivity to Anticancer Drugs of Patients with a Pancreatic Adenocarcinoma. The American Journal of Pathology 185:1022-1032, 2015

111. Buscail L, Bournet B, Cordelier P: Role of oncogenic KRAS in the diagnosis, prognosis and treatment of pancreatic cancer. Nature Reviews Gastroenterology & Hepatology 17:153-168, 2020

112. Ogura T, Yamao K, Hara K, et al: Prognostic value of K-ras mutation status and subtypes in endoscopic ultrasound-guided fine-needle aspiration specimens from patients with unresectable pancreatic cancer. J Gastroenterol 48:640-6, 2013

113. Le N, Sund M, Vinci A: Prognostic and predictive markers in pancreatic adenocarcinoma. Digestive and Liver Disease 48:223-230, 2016

114. Collisson EA, Bailey P, Chang DK, et al: Molecular subtypes of pancreatic cancer. Nature Reviews Gastroenterology & Hepatology 16:207-220, 2019

115. Wang J-D, Jin K, Chen X-Y, et al: Clinicopathological significance of SMAD4 loss in pancreatic ductal adenocarcinomas: a systematic review and meta-analysis. Oncotarget 8, 2016

116. Ormanns S, Haas M, Remold A, et al: The Impact of SMAD4 Loss on Outcome in Patients with Advanced Pancreatic Cancer Treated with Systemic Chemotherapy. International Journal of Molecular Sciences 18:1094, 2017

117. Yamada S, Fujii T, Shimoyama Y, et al: SMAD4 expression predicts local spread and treatment failure in resected pancreatic cancer. Pancreas 44:660-4, 2015

118. Biankin AV, Kench JG, Colvin EK, et al: Expression of S100A2 calcium-binding protein predicts response to pancreatectomy for pancreatic cancer. Gastroenterology 137:558-68, 568.e1-11, 2009

119. Bachet J-B, Maréchal R, Demetter P, et al: S100A2 is a predictive biomarker of adjuvant therapy benefit in pancreatic adenocarcinoma. European Journal of Cancer 49:2643-2653, 2013

120. Dreyer SB, Pinese M, Jamieson NB, et al: Precision Oncology in Surgery: Patient Selection for Operable Pancreatic Cancer. Annals of Surgery 272, 2020

121. Marcus L, Lemery SJ, Keegan P, et al: FDA Approval Summary: Pembrolizumab for the Treatment of Microsatellite Instability-High Solid Tumors. Clinical Cancer Research 25:3753-3758, 2019

122. Ormanns S, Heinemann V, Raponi M, et al: Human equilibrative nucleoside transporter 1 is not predictive for gemcitabine efficacy in advanced pancreatic cancer: translational results from the AIO-PK0104 phase III study with the clone SP120 rabbit antibody. Eur J Cancer 50:1891-9, 2014

123. Boeck S, Jung A, Laubender RP, et al: EGFR pathway biomarkers in erlotinib-treated patients with advanced pancreatic cancer: translational results from the randomised, crossover phase 3 trial AIO-PK0104. British Journal of Cancer 108:469-476, 2013

124. Philip PA, Lutz MP: Targeting Epidermal Growth Factor Receptor-Related Signaling Pathways in Pancreatic Cancer. Pancreas 44:1046-52, 2015

125. Barton CM, Hall PA, Hughes CM, et al: Transforming growth factor alpha and epidermal growth factor in human pancreatic cancer. J Pathol 163:111-6, 1991

126. Valsecchi ME, McDonald M, Brody JR, et al: Epidermal growth factor receptor and insulinlike growth factor 1 receptor expression predict poor survival in pancreatic ductal adenocarcinoma. Cancer 118:3484-93, 2012

127. Chiramel J, Backen AC, Pihlak R, et al: Targeting the Epidermal Growth Factor Receptor in Addition to Chemotherapy in Patients with Advanced Pancreatic Cancer: A Systematic Review and Meta-Analysis. Int J Mol Sci 18, 2017

128. Yoon S, Seger R: The extracellular signal-regulated kinase: multiple substrates regulate diverse cellular functions. Growth Factors 24:21-44, 2006

129. Sahin IH, Iacobuzio-Donahue CA, O'Reilly EM: Molecular signature of pancreatic adenocarcinoma: an insight from genotype to phenotype and challenges for targeted therapy. Expert Opin Ther Targets 20:341-59, 2016

130. Minden A, Lin A, McMahon M, et al: Differential activation of ERK and JNK mitogen-activated protein kinases by Raf-1 and MEKK. Science 266:1719-23, 1994

131. Suzuki S, Okada M, Shibuya K, et al: JNK suppression of chemotherapeutic agents-induced ROS confers chemoresistance on pancreatic cancer stem cells. Oncotarget 6:458-470, 2015

132. Yuan XP, Dong M, Li X, et al: GRP78 promotes the invasion of pancreatic cancer cells by FAK and JNK. Mol Cell Biochem 398:55-62, 2015

133. Song M, Bode AM, Dong Z, et al: AKT as a Therapeutic Target for Cancer. Cancer Res 79:1019-1031, 2019

134. Baer R, Cintas C, Therville N, et al: Implication of PI3K/Akt pathway in pancreatic cancer: When PI3K isoforms matter? Adv Biol Regul 59:19-35, 2015

135. Ruggeri BA, Huang L, Wood M, et al: Amplification and overexpression of the AKT2 oncogene in a subset of human pancreatic ductal adenocarcinomas. Mol Carcinog 21:81-6, 1998

136. Murthy D, Attri KS, Singh PK: Phosphoinositide 3-Kinase Signaling Pathway in Pancreatic Ductal Adenocarcinoma Progression, Pathogenesis, and Therapeutics. Frontiers in physiology 9:335-335, 2018

137. Morton JP, Mongeau ME, Klimstra DS, et al: Sonic hedgehog acts at multiple stages during pancreatic tumorigenesis. Proceedings of the National Academy of Sciences 104:5103-5108, 2007

138. Derynck R, Zhang YE: Smad-dependent and Smad-independent pathways in TGF-beta family signalling. Nature 425:577-84, 2003

139. Bierie B, Moses HL: TGFβ: the molecular Jekyll and Hyde of cancer. Nature Reviews Cancer 6:506-520, 2006

140. Jones S, Zhang X, Parsons DW, et al: Core Signaling Pathways in Human Pancreatic Cancers Revealed by Global Genomic Analyses. Science 321:1801-1806, 2008

141. Lin X, Feng XH: Abrogation of transforming growth factor-beta signaling in pancreatic cancer. World J Surg 29:312-6, 2005

142. Sinn M, Bahra M, Liersch T, et al: CONKO-005: Adjuvant Chemotherapy With Gemcitabine Plus Erlotinib Versus Gemcitabine Alone in Patients After R0 Resection of Pancreatic Cancer: A Multicenter Randomized Phase III Trial. Journal of Clinical Oncology:JCO.2017.72.6463, 2017

143. Gundry Michael C, Brunetti L, Lin A, et al: Highly Efficient Genome Editing of Murine and Human Hematopoietic Progenitor Cells by CRISPR/Cas9. Cell Reports 17:1453-1461, 2016

144. Au - Brunetti L, Au - Gundry MC, Au - Kitano A, et al: Highly Efficient Gene Disruption of Murine and Human Hematopoietic Progenitor Cells by CRISPR/Cas9. JoVE:e57278, 2018

145. Benjamini Y, Hochberg Y: Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B (Methodological) 57:289-300, 1995

146. Synthego: Synthego Performance Analysis, ICE Analysis (ed v2.0), Synthego, 2019

147. Vincent A, Herman J, Schulick R, et al: Pancreatic cancer. Lancet (London, England) 378:607-620, 2011

148. Sausen M, Phallen J, Adleff V, et al: Clinical implications of genomic alterations in the tumour and circulation of pancreatic cancer patients. Nat Commun, 2015

149. Kosmidis C, Sapalidis K, Kotidis E, et al: Pancreatic cancer from bench to bedside: molecular pathways and treatment options. Annals of Translational Medicine 4, 2016

150. GeneDx: Pancreatic Cancer Panel. https://www.genedx.com/test-catalog/available-tests/pancreatic-cancer-panel/, 2016

151. Seki M, Kimura S, Isobe T, et al: Recurrent SPI1 (PU.1) fusions in high-risk pediatric T cell acute lymphoblastic leukemia. Nature Genetics 49:1274, 2017

152. Shiraishi Y, Sato Y, Chiba K, et al: An empirical Bayesian framework for somatic mutation detection from cancer genome sequencing data. Nucleic Acids Research 41:e89-e89, 2013

153. Robinson JT, Thorvaldsdóttir H, Winckler W, et al: Integrative genomics viewer. Nature Biotechnology 29:24-26, 2011

154. Yoshizato T, Nannya Y, Atsuta Y, et al: Genetic abnormalities in myelodysplasia and secondary acute myeloid leukemia: impact on outcome of stem cell transplantation. Blood 129:2347-2358, 2017

155. Hüebschmann D, Kurzawa N, Steinhauser S, et al: Deciphering programs of transcriptional regulation by combined deconvolution of multiple omics layers. bioRxiv:199547, 2017

156. Henke M, Mattern D, Pepe M, et al: Do erythropoietin receptors on cancer cells explain unexpected clinical findings? J Clin Oncol 24:4708-13, 2006

157. Takai E, Totoki Y, Nakamura H, et al: Clinical utility of circulating tumor DNA for molecular assessment in pancreatic cancer. Scientific Reports 5:18425, 2015

158. AmbryGenetics: Pancreatic Cancer Panel. http://www.ambrygen.com/tests/pancnext, 2016

159. Invitae: Pancreatic Cancer Panel. https://www.invitae.com/en/physician/tests/01261/, 2016

160. Kanehisa M, Goto S: KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28:27-30, 2000

161. Karandish F, Mallik S: Biomarkers and Targeted Therapy in Pancreatic Cancer. Biomarkers in Cancer:27-35, 2016

162. Beroukhim R, Mermel CH, Porter D, et al: The landscape of somatic copy-number alteration across human cancers. Nature 463:899-905, 2010

163. Shlien A, Malkin D: Copy number variations and cancer. Genome Medicine 1:62-62, 2009

164. Taylor BS, Barretina J, Socci ND, et al: Functional Copy-Number Alterations in Cancer. PLoS ONE 3:e3179, 2008

165. Zack TI, Schumacher SE, Carter SL, et al: Pan-cancer patterns of somatic copy number alteration. Nat Genet 45:1134-1140, 2013

166. Christen F, Hoyer K, Yoshida K, et al: Genomic landscape and clonal evolution of acute myeloid leukemia with t(8;21): an international study on 331 patients. Blood 133:1140-1151, 2019

167. Gao J, Aksoy BA, Dogrusoz U, et al: Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal 6:pl1, 2013

168. Cerami E, Gao J, Dogrusoz U, et al: The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. Cancer Discovery 2:401-404, 2012

169. Tempero MA, Uchida E, Takasaki H, et al: Relationship of carbohydrate antigen 19-9 and Lewis antigens in pancreatic cancer. Cancer Res 47:5501-3, 1987

170. Uhlen M, Zhang C, Lee S, et al: A pathology atlas of the human cancer transcriptome. Science 357:eaan2507, 2017

171. Kruger S, Haas M, Burger PJ, et al: Isolated pulmonary metastases define a favorable subgroup in metastatic pancreatic cancer. Pancreatology 16:593-8, 2016

172. Bailey P, Chang DK, Nones K, et al: Genomic analyses identify molecular subtypes of pancreatic cancer. Nature 531:47-52, 2016

173. Alexandrov Ludmil B, Nik-Zainal S, Wedge David C, et al: Deciphering Signatures of Mutational Processes Operative in Human Cancer. Cell Reports 3:246-259, 2013

174. Chapuy B, Stewart C, Dunford AJ, et al: Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. Nature Medicine 24:679-690, 2018

175. Ma J, Setton J, Lee NY, et al: The therapeutic significance of mutational signatures from DNA repair deficiency in cancer. Nature Communications 9:3292, 2018

176. Alexandrov LB, Ju YS, Haase K, et al: Mutational signatures associated with tobacco smoking in human cancer. Science (New York, N.Y.) 354:618-622, 2016

177. Manole S, Richards EJ, Meyer AS: JNK Pathway Activation Modulates Acquired Resistance to EGFR/HER2-Targeted Therapies. Cancer research 76:5219-5228, 2016

178. Bardeesy N, Cheng KH, Berger JH, et al: Smad4 is dispensable for normal pancreas development yet critical in progression and tumor biology of pancreas cancer. Genes Dev 20:3130-46, 2006

179. Izeradjene K, Combs C, Best M, et al: Kras(G12D) and Smad4/Dpc4 haploinsufficiency cooperate to induce mucinous cystic neoplasms and invasive adenocarcinoma of the pancreas. Cancer Cell 11:229-43, 2007

180. Fryer RA, Barlett B, Galustian C, et al: Mechanisms underlying gemcitabine resistance in pancreatic cancer and sensitisation by the iMiD lenalidomide. Anticancer Res 31:3747-56, 2011

181. Ho J, Li X, Zhang L, et al: Translational genomics in pancreatic ductal adenocarcinoma: A review with re-analysis of TCGA dataset. Seminars in Cancer Biology, 2018

182. Singh RR, O'Reilly EM: New Treatment Strategies for Metastatic Pancreatic Ductal Adenocarcinoma. Drugs 80:647-669, 2020

183. Witkiewicz AK, McMillan EA, Balaji U, et al: Whole-exome sequencing of pancreatic cancer defines genetic diversity and therapeutic targets. Nat Commun 6, 2015

184. Huang L, Yu D, Wu C, et al: Copy number variation at 6q13 functions as a long-range regulator and is associated with pancreatic cancer risk. Carcinogenesis 33:94-100, 2011

185. Mueller S, Engleitner T, Maresch R, et al: Evolutionary routes and KRAS dosage define pancreatic cancer phenotypes. Nature 554:62-68, 2018

186. Aung KL, Fischer SE, Denroche RE, et al: Genomics-Driven Precision Medicine for Advanced Pancreatic Cancer: Early Results from the COMPASS Trial. Clin Cancer Res 24:1344-1354, 2018

187. Chan-Seng-Yue M, Kim JC, Wilson GW, et al: Transcription phenotypes of pancreatic cancer are driven by genomic events during tumor evolution. Nature Genetics 52:231-240, 2020

188. Bernard V, Semaan A, Huang J, et al: Single-Cell Transcriptomics of Pancreatic Cancer Precursors Demonstrates Epithelial and Microenvironmental Heterogeneity as an Early Event in Neoplastic Progression. Clinical Cancer Research 25:2194-2205, 2019

189. Peng J, Sun B-F, Chen C-Y, et al: Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma. Cell Research 29:725-738, 2019

190. Lomberk G, Blum Y, Nicolle R, et al: Distinct epigenetic landscapes underlie the pathobiology of pancreatic cancer subtypes. Nature Communications 9:1978, 2018

191. Zhou W, Chen L, Li C, et al: The multifaceted roles of long noncoding RNAs in pancreatic cancer: an update on what we know. Cancer Cell International 20:41, 2020

192. Pan S, Brentnall TA, Chen R: Proteome alterations in pancreatic ductal adenocarcinoma. Cancer Lett 469:429-436, 2020

193. Yap TA, Yan L, Patnaik A, et al: First-in-man clinical trial of the oral pan-AKT inhibitor MK-2206 in patients with advanced solid tumors. J Clin Oncol 29:4688-95, 2011

194. Golan T, Hammel P, Reni M, et al: Maintenance Olaparib for Germline BRCA-Mutated Metastatic Pancreatic Cancer. N Engl J Med 381:317-327, 2019

195. Samowitz WS, Sweeney C, Herrick J, et al: Poor survival associated with the BRAF V600E mutation in microsatellite-stable colon cancers. Cancer Res 65:6063-9, 2005

196. Onitilo AA, Engel JM, Greenlee RT, et al: Breast cancer subtypes based on ER/PR and Her2 expression: comparison of clinicopathologic features and survival. Clinical medicine & research 7:4-13, 2009

197. Paik S, Tang G, Shak S, et al: Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. J Clin Oncol 24:3726-34, 2006

198. Ballehaninna UK, Chamberlain RS: The clinical utility of serum CA 19-9 in the diagnosis, prognosis and management of pancreatic adenocarcinoma: An evidence based appraisal. Journal of gastrointestinal oncology 3:105-119, 2012

199. Kaneda MM, Cappello P, Nguyen AV, et al: Macrophage PI3Kγ Drives Pancreatic Ductal Adenocarcinoma Progression. Cancer Discovery 6:870-885, 2016

200. Brennan MF, Kattan MW, Klimstra D, et al: Prognostic nomogram for patients undergoing resection for adenocarcinoma of the pancreas. Ann Surg 240:293-8, 2004

201. Vernerey D, Huguet F, Vienot A, et al: Prognostic nomogram and score to predict overall survival in locally advanced untreated pancreatic cancer (PROLAP). Br J Cancer 115:281-9, 2016

202. Hang J, Wu L, Zhu L, et al: Prediction of overall survival for metastatic pancreatic cancer: Development and validation of a prognostic nomogram with data from open clinical trial and real-world study. Cancer Med 7:2974-84, 2018

203. Dell'Aquila E, Fulgenzi CAM, Minelli A, et al: Prognostic and predictive factors in pancreatic cancer. Oncotarget 11:924-941, 2020

204. Steins M, Thomas M, Geißler M: Erlotinib. Recent Results Cancer Res 211:1-17, 2018

205. Conradt L, Godl K, Schaab C, et al: Disclosure of erlotinib as a multikinase inhibitor in pancreatic ductal adenocarcinoma. Neoplasia (New York, N.Y.) 13:1026-1034, 2011

206. Bareschino MA, Schettino C, Troiani T, et al: Erlotinib in cancer treatment. Annals of Oncology 18:vi35-vi41, 2007

207. Shugang X, Hongfa Y, Jianpeng L, et al: Prognostic Value of SMAD4 in Pancreatic Cancer: A Meta-Analysis. Translational oncology 9:1-7, 2016

208. Crane CH, Varadhachary GR, Yordy JS, et al: Phase II trial of cetuximab, gemcitabine, and oxaliplatin followed by chemoradiation with cetuximab for locally advanced (T4) pancreatic adenocarcinoma: correlation of Smad4(Dpc4) immunostaining with pattern of disease progression. J Clin Oncol 29:3037-43, 2011

209. Herman JM, Fan KY, Wild AT, et al: Correlation of Smad4 status with outcomes in patients receiving erlotinib combined with adjuvant chemoradiation and chemotherapy after resection for pancreatic adenocarcinoma. International journal of radiation oncology, biology, physics 87:458-459, 2013

210. Derynck R, Goeddel DV, Ullrich A, et al: Synthesis of Messenger RNAs for Transforming Growth Factors α and β and the Epidermal Growth Factor Receptor by Human Tumors. Cancer Research 47:707-712, 1987

211. Shen W, Tao G-q, Zhang Y, et al: TGF-β in pancreatic cancer initiation and progression: two sides of the same coin. Cell & Bioscience 7:39, 2017

212. Duda DG, Sunamura M, Lefter LP, et al: Restoration of SMAD4 by gene therapy reverses the invasive phenotype in pancreatic adenocarcinoma cells. Oncogene 22:6857-64, 2003

213. Evan GI, Vousden KH: Proliferation, cell cycle and apoptosis in cancer. Nature 411:342-8, 2001

214. Furukawa T, Sunamura M, Horii A: Molecular mechanisms of pancreatic carcinogenesis. Cancer Sci 97:1-7, 2006

215. David Charles J, Huang Y-H, Chen M, et al: TGF-beta Tumor Suppression through a Lethal EMT. Cell 164:1015-1030, 2016

216. Ahmed S, Bradshaw A-D, Gera S, et al: The TGF-β/Smad4 Signaling Pathway in Pancreatic Carcinogenesis and Its Clinical Significance. Journal of clinical medicine 6:5, 2017

217. Bedi A, Chang X, Noonan K, et al: Inhibition of TGF-β Enhances the In Vivo Antitumor Efficacy of EGF Receptor–Targeted Therapy. Molecular Cancer Therapeutics 11:2429-2439, 2012

218. Ozawa H, Ranaweera RS, Izumchenko E, et al: SMAD4 Loss Is Associated with Cetuximab Resistance and Induction of MAPK/JNK Activation in Head and Neck Cancer Cells. Clinical Cancer Research 23:5162-5175, 2017

219. Chen Y-W, Hsiao P-J, Weng C-C, et al: SMAD4 loss triggers the phenotypic changes of pancreatic ductal adenocarcinoma cells. BMC cancer 14:181-181, 2014

220. Hsieh Y-Y, Liu T-P, Chou C-J, et al: Integration of Bioinformatics Resources Reveals the Therapeutic Benefits of Gemcitabine and Cell Cycle Intervention in SMAD4-Deleted Pancreatic Ductal Adenocarcinoma. Genes 10:766, 2019

221. Cui Y, Brosnan JA, Blackford AL, et al: Genetically defined subsets of human pancreatic cancer show unique in vitro chemosensitivity. Clin Cancer Res 18:6519-30, 2012

222. Shin SH, Kim HJ, Hwang DW, et al: The DPC4/SMAD4 genetic status determines recurrence patterns and treatment outcomes in resected pancreatic ductal adenocarcinoma: A prospective cohort study. Oncotarget 8:17945-17959, 2017

223. Herman JM, Jabbour SK, Lin SH, et al: Smad4 Loss Correlates With Higher Rates of Local and Distant Failure in Pancreatic Adenocarcinoma Patients Receiving Adjuvant Chemoradiation. Pancreas 47:208-212, 2018

224. Polyak K, Weinberg RA: Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. Nat Rev Cancer 9:265-73, 2009

225. Dardare J, Witz A, Merlin JL, et al: SMAD4 and the TGFβ Pathway in Patients with Pancreatic Ductal Adenocarcinoma. Int J Mol Sci 21, 2020

226. Zhang X, Cao J, Pei Y, et al: Smad4 inhibits cell migration via suppression of JNK activity in human pancreatic carcinoma PANC-1 cells. Oncology letters 11:3465-3470, 2016

227. Huang W, Navarro-Serer B, Jeong YJ, et al: Pattern of Invasion in Human Pancreatic Cancer Organoids Is Associated with Loss of SMAD4 and Clinical Outcome. Cancer Research 80:2804-2817, 2020

228. Grimminger PP, Stephens C, Waldschmidt D, et al: C-kit mRNA expression in pancreatic adenocarcinoma and matched stromal tissue: Prognostic and therapeutic implications. Journal of Clinical Oncology 32:e15185-e15185, 2014

229. Tanaka M, Suzuki HI, Shibahara J, et al: EVI1 oncogene promotes KRAS pathway through suppression of microRNA-96 in pancreatic carcinogenesis. Oncogene 33:2454-63, 2014

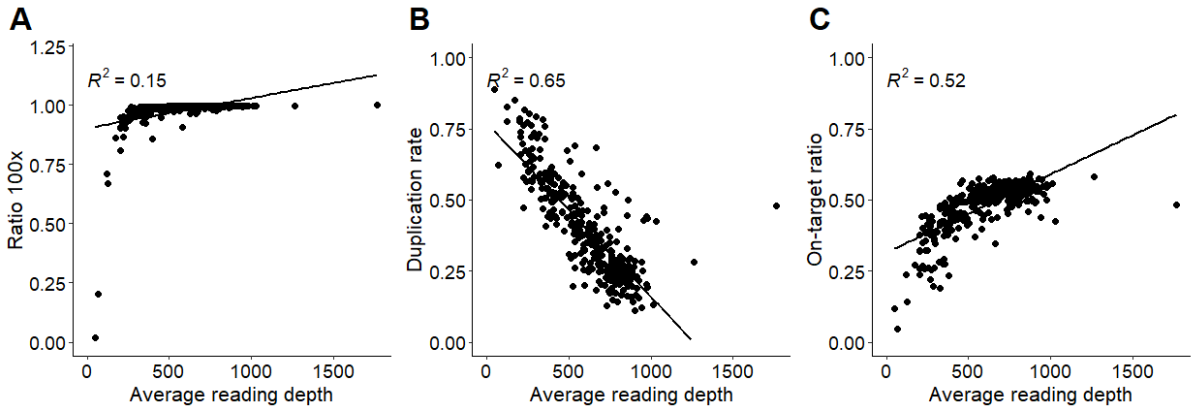
230. Li J, Yuan S, Norgard RJ, et al: Epigenetic and transcriptional control of the epidermal growth factor receptor (EGFR) regulates the tumor immune microenvironment in pancreatic cancer. Cancer Discovery:CD-20-0519, 2020

231. Balli D, Rech AJ, Stanger BZ, et al: Immune Cytolytic Activity Stratifies Molecular Subsets of Human Pancreatic Cancer. Clinical Cancer Research 23:3129-3138, 2017

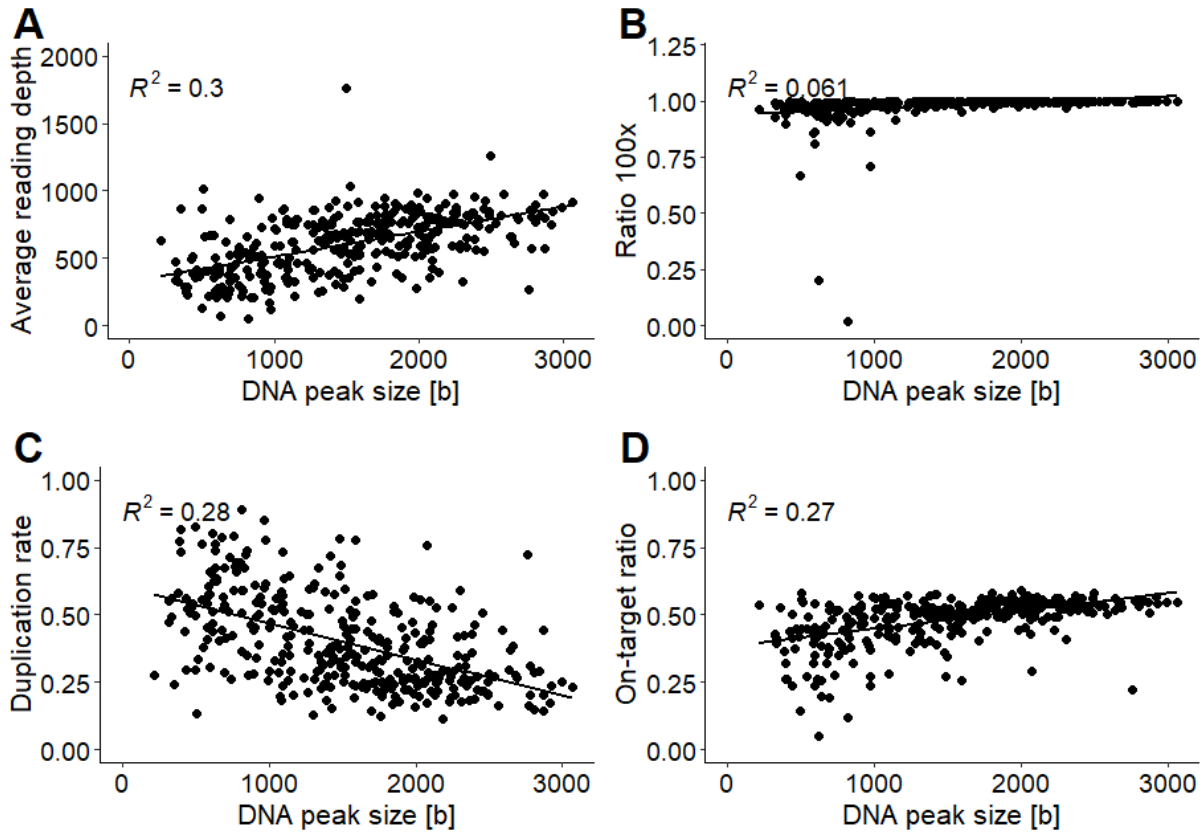
232. Muller Patricia AJ, Vousden Karen H: Mutant p53 in Cancer: New Functions and Therapeutic Opportunities. Cancer Cell 25:304-317, 2014

Appendix

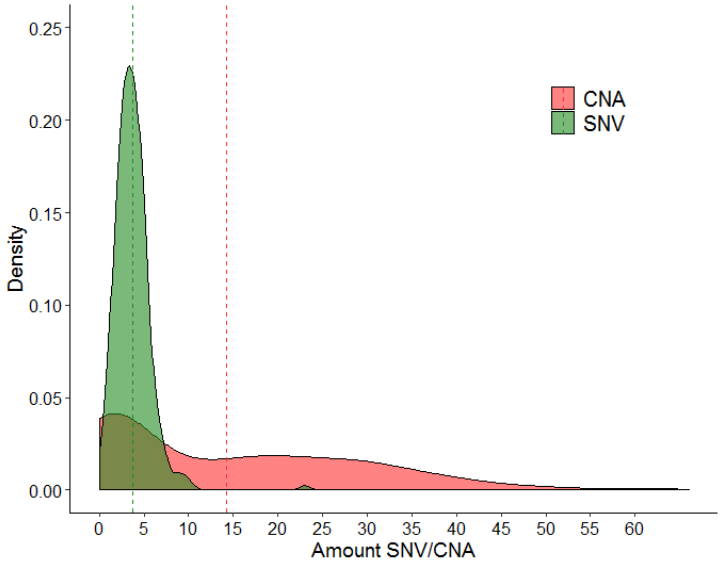
Supplemental Figures



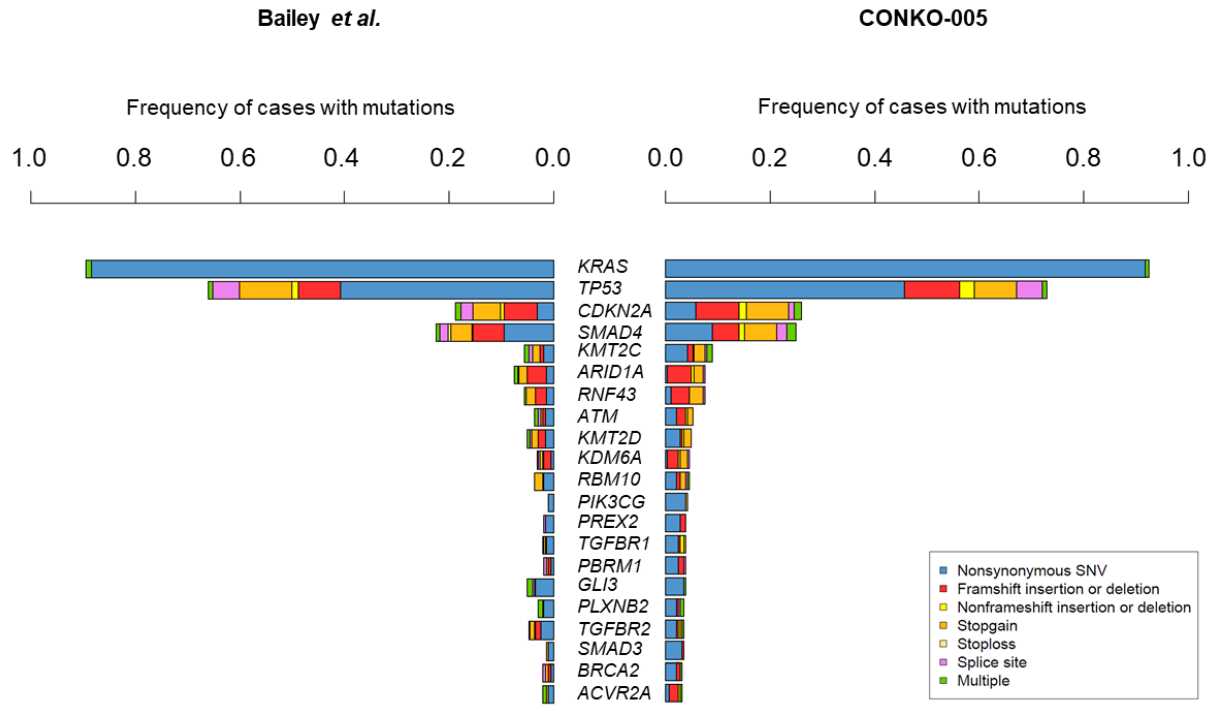
Supplemental Figure 1: Correlation of QC criteria with Average reading depth. A direct correlation can be found between average reading depth and **A)** Ratio 100x **B)** duplication rate or **C)** on-target ratio, as seen by the high proportion of variation in the dependent (response) variable (R^2) that is explained by the linear regression model.



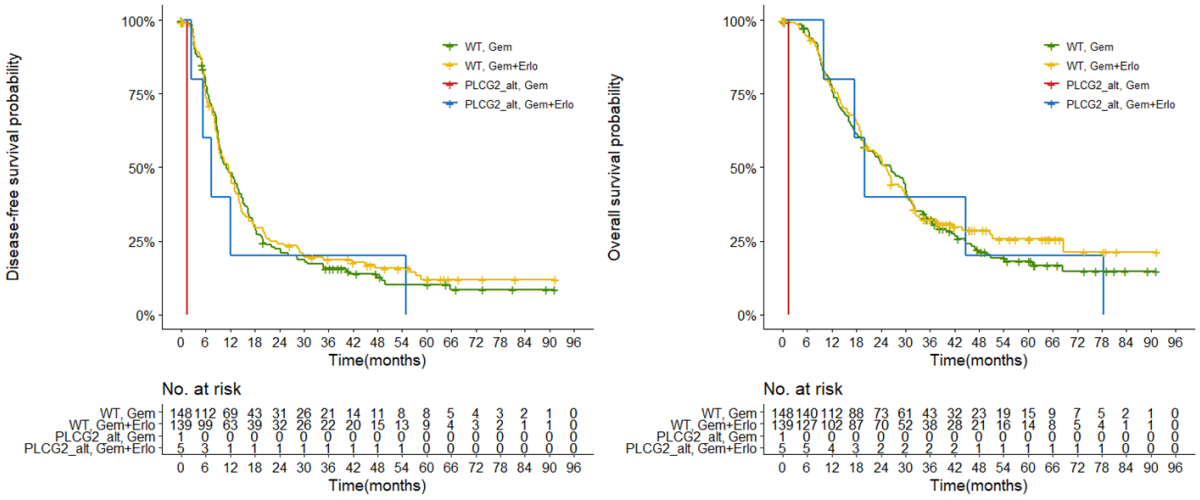
Supplemental Figure 2: Correlation of QC criteria with DNA peak size. No correlation can be found between peak size of genomic DNA and **A)** average reading depth **B)** Ratio 100x, **C)** duplication rate or **D)** on-target ratio, as seen by the low proportion of variation in the dependent (response) variable (R^2) that is explained by the linear regression model.



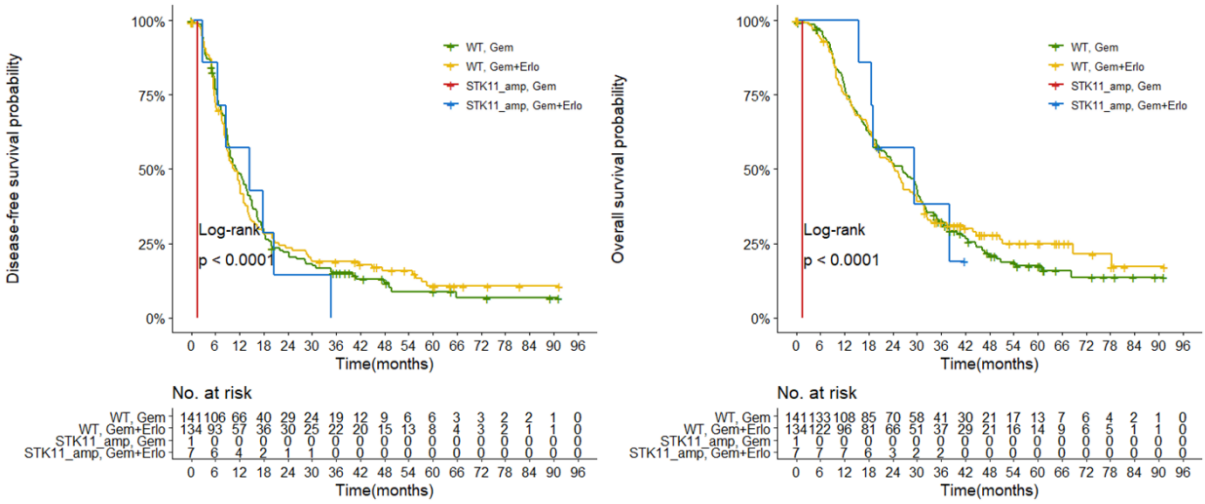
Supplemental Figure 3: Distribution of SNV/CNA per patient in complete cohort. Dashed lines mark the median alteration amount. Previously published in Hoyer et al.[1]



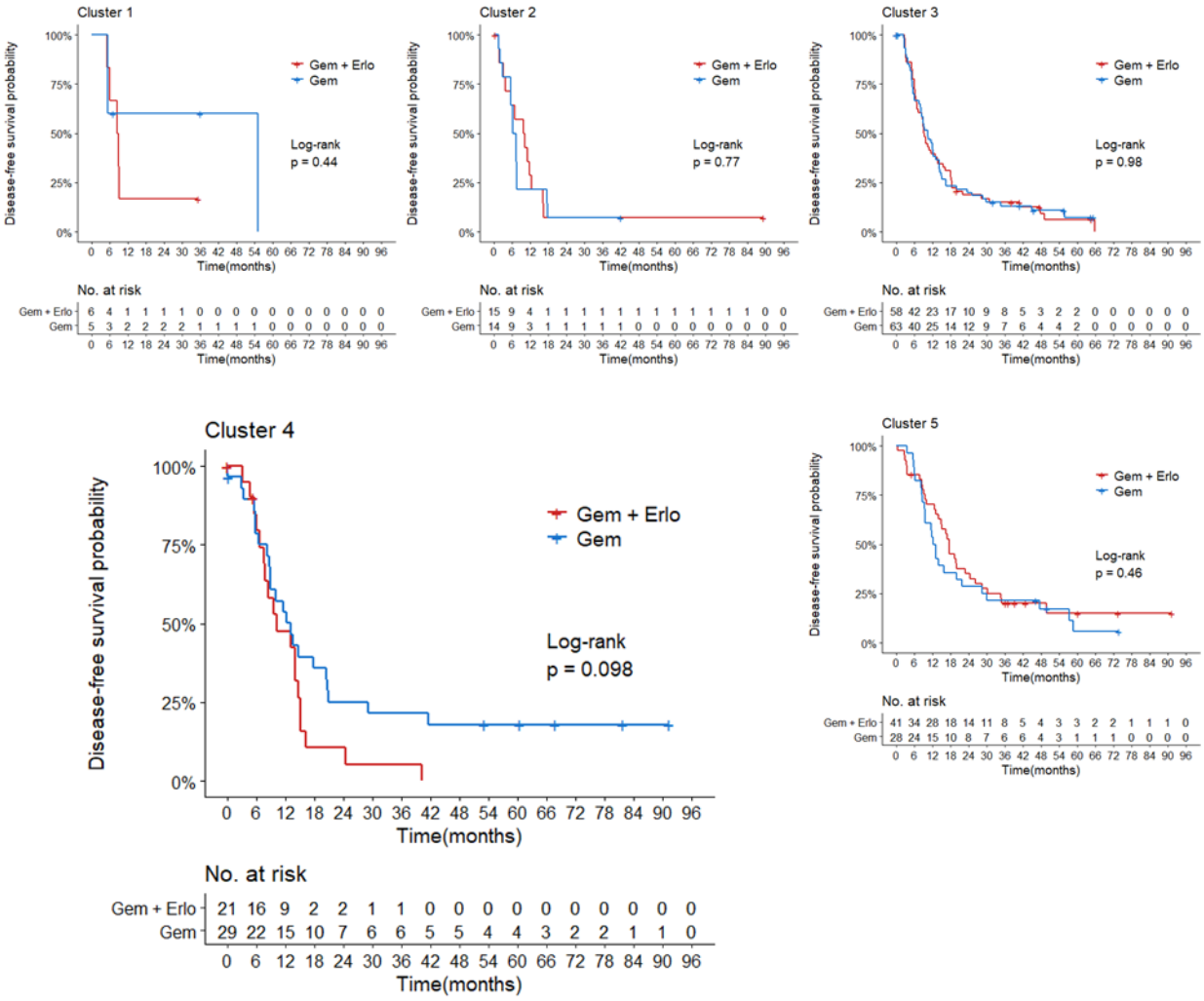
Supplemental Figure 4: Comparison of mutation frequencies and type with findings from Bailey et al.3 The 20 most frequently altered genes are shown in order of mutation frequency, based on the CONKO-005 cohort results. Mutations types are indicated by colour. Modified from Hoyer et al.[1]



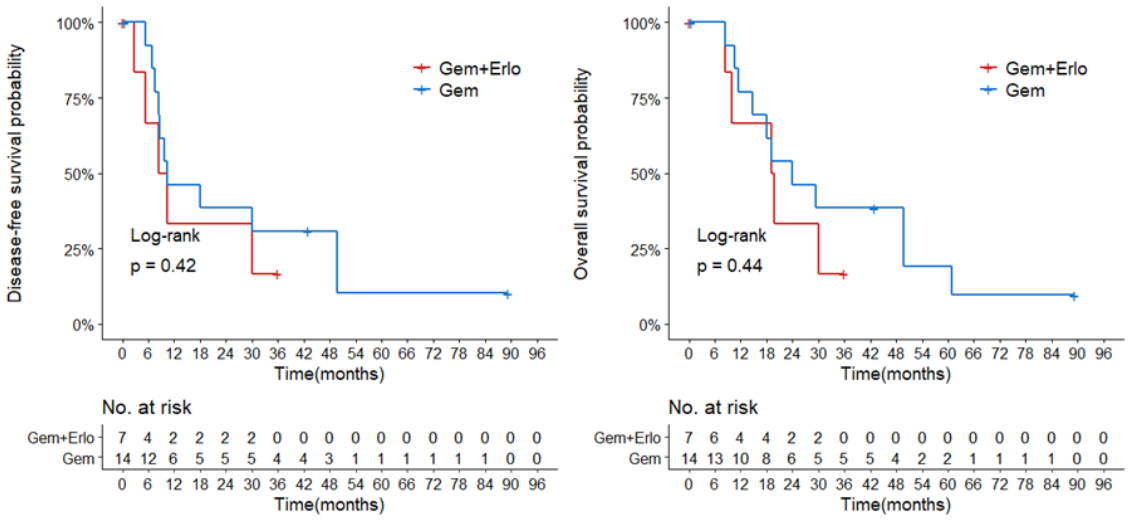
Supplemental Figure 5: Erlotinib sensitivity in *PLCG2* altered PDAC patients. Kaplan-Meier curve comparing OS between both treatment arms in *PLCG2* altered (PLCG2_alt) and *PLCG2* wild-type (WT) patients. Log rank p-values are <0.001 for both.



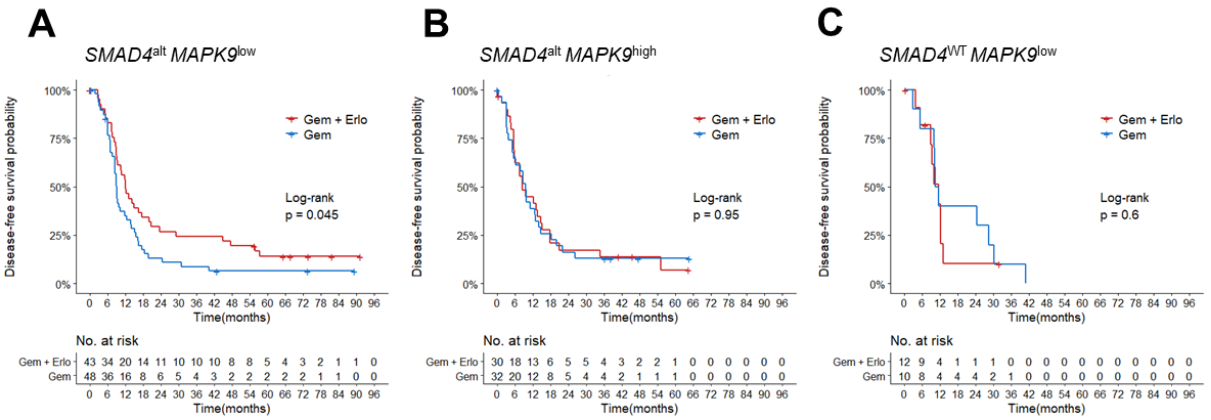
Supplemental Figure 6: Erlotinib sensitivity in *STK11* amplified PDAC patients. Kaplan-Meier curve comparing OS between both treatment arms in *STK11* amplified (STK11_amp) and *STK11* wild-type (WT) patients. Log rank p-values are <0.001 for both.



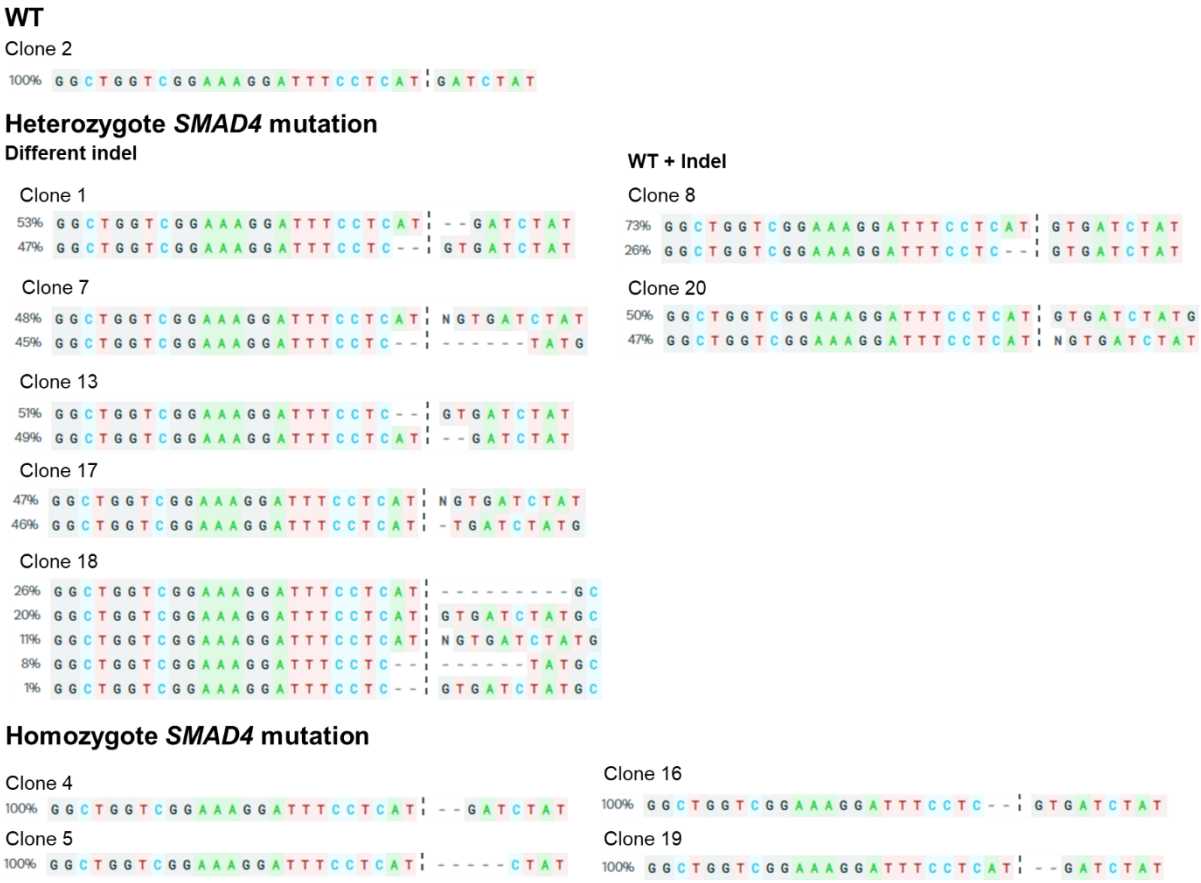
Supplemental Figure 7: KM of DFS comparing both treatment arms for each cluster. All patients in each cluster are included and divided depending on their treatment arm. Gem = Gemcitabine, Gem + Erlo = Gemcitabine and erlotinib combination therapy.



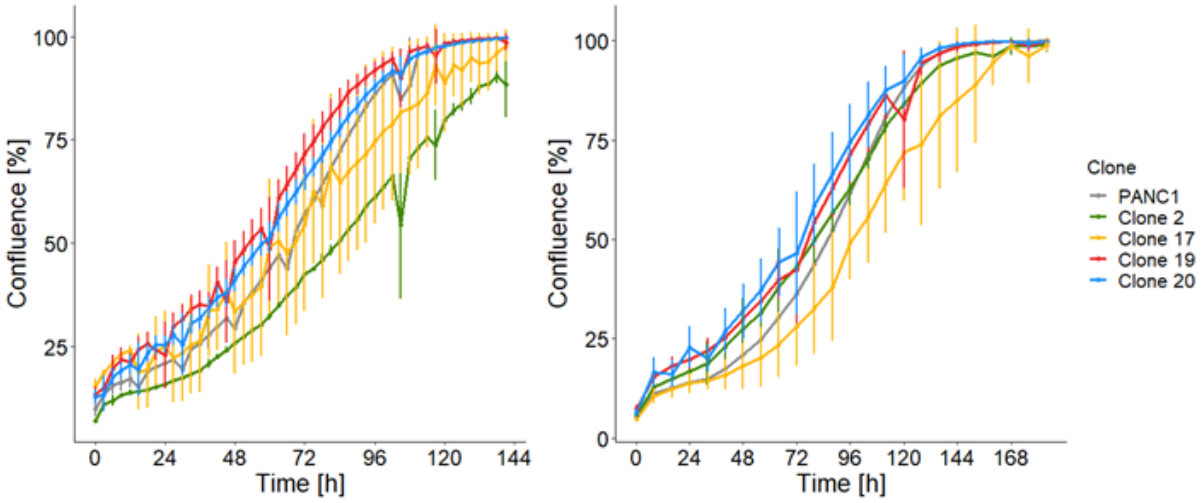
Supplemental Figure 8: KM curves of treatment arms in *EGFR* altered patients. Curves compare treatment arms in all patients with either *EGFR* mutation (n = 4) or amplification (n = 17). Modified from Hoyer et al.[1]



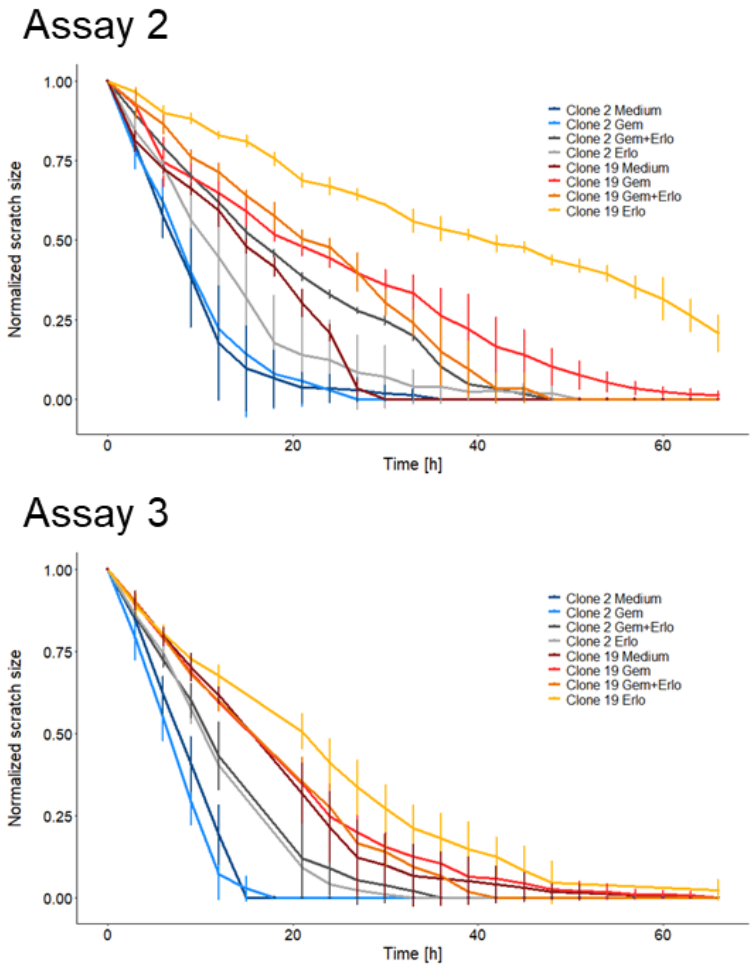
Supplemental Figure 9: DFS Kaplan-Meier curves of treatment arms within patient groups with different *SMAD4 MAPK9* status. A) *SMAD4*^{alt} *MAPK9*^{low} patients live significantly longer when treated with erlotinib, while neither **B)** *SMAD4*^{alt} *MAPK9*^{high}, nor **C)** *SMAD4*^{WT} *MAPK9*^{low} patients exhibit that effect.



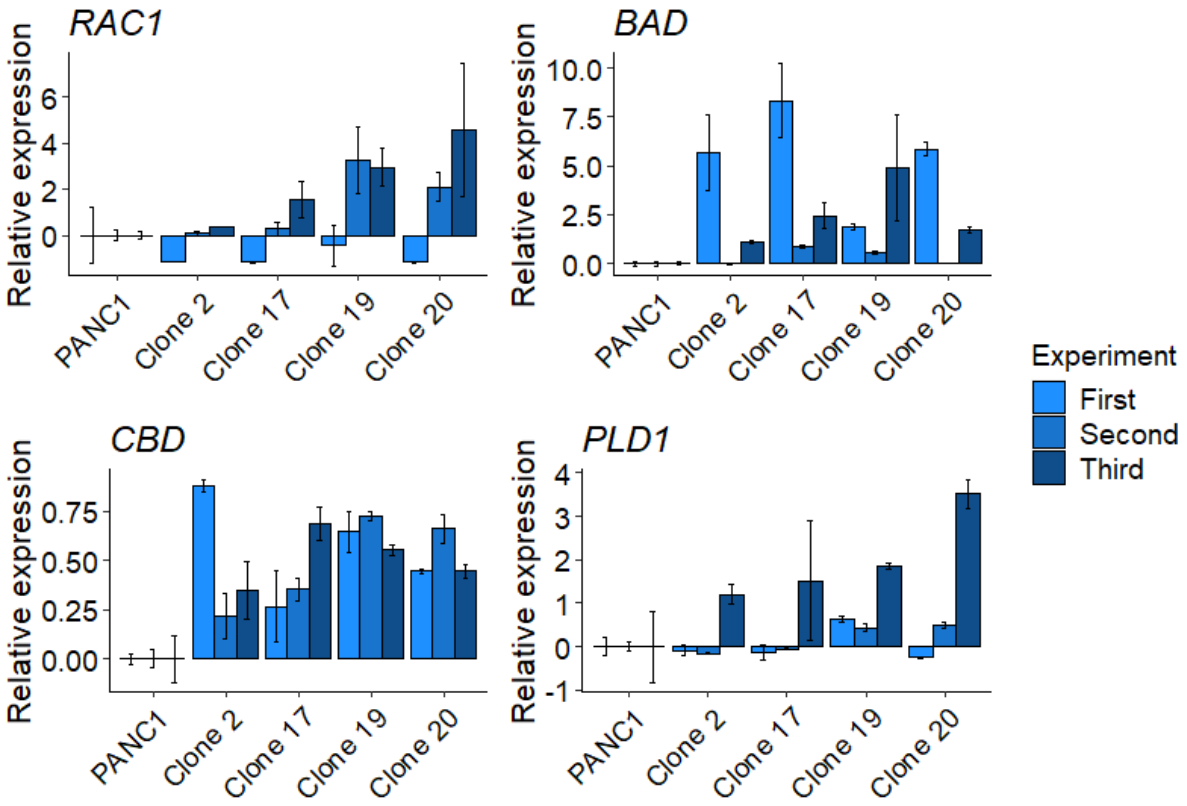
Supplemental Figure 10: Overview of the twelve CRIPR/Cas9 SMAD4 KO SSC. Deconstructed Sanger sequencing tracks using the ICE program to calculate each variants share of the complete Sanger track. Normalized relative contribution of each sequence found in the bulk one week after transfection is shown. The contributions show the inferred sequences present in the edited population and their relative proportions (in contrast to the Indel plot that does not specify sequence contributions). Cut sites are represented by black vertical dotted lines. [Modified from https://ice.synthego.com/]



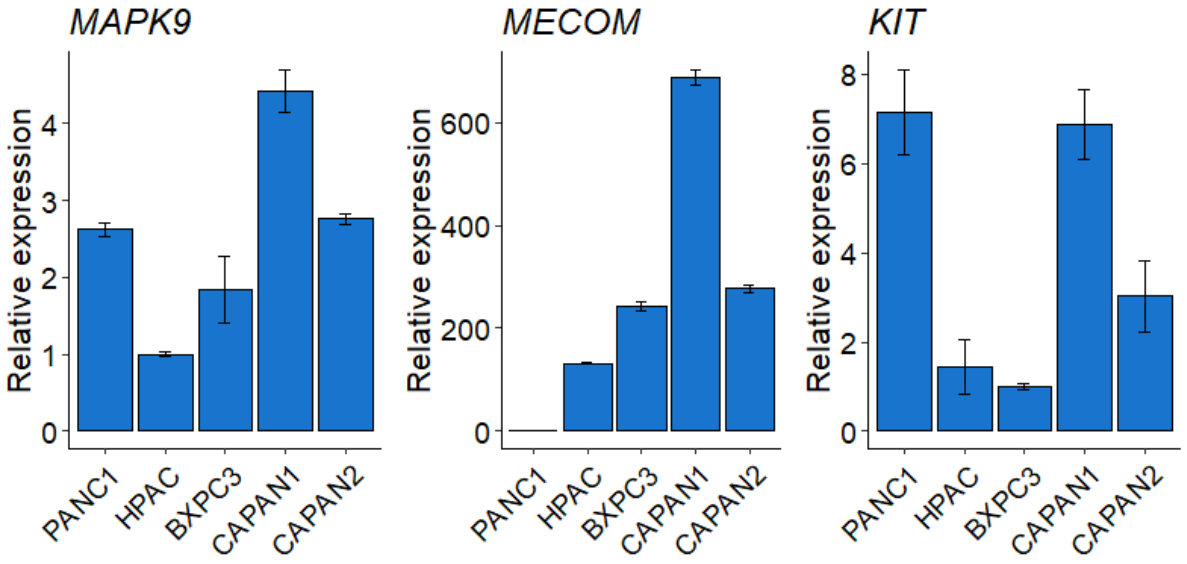
Supplemental Figure 11: Repeats of the SSC proliferation measurements. No significant difference in cell proliferation could be seen between the four clones and the cell line.



Supplemental Figure 12: Cell migration of *SMAD4* KO single cell clones with or without treatment. The results of the wound scratch assay comparing the WT clone (clone 2) with the SMAD4 KO clone (clone 19). For each clone $1.5x10^5$ cells were incubated in different medium conditions. After applying the scratch, cells were continuously observed for four days. Pictures were taken every three hours in triplicates for each condition. Lines and error bars show mean ± standard deviation of triplicates.



Supplemental Figure 13: Expression of the other genes that were found to be differentially expressed between *SMAD4*^{alt} and *SMAD4* WT patients in *SMAD4* KO single cell clones. Relative expression of the four genes that were found to be significantly increased or decreased in *SMAD4*^{alt} patients but do not have a different expression in SMAD4 KO cells. ΔΔCT values normalized based on *GAPDH* CT and ΔCT values of PANC1.



Supplemental Figure 14: Expression of MAPK9, MECOM and KIT in PDAC cell lines.

Supplemental Tables

Supplemental Table 1: Morphological estimation of tumour content of all 293 samples. Previously published in Hoyer et al.[1]

Patient ID	Tumor Content	Patient ID	Tumor Content	Patient ID	Tumor Content	Patient ID	Tumor Content	Patient ID	Tumor Content
ID001	0.4	ID061	0.1	ID121	0.1	ID181	0.1	ID241	0.15
ID002	0.5	ID062	0.3	ID122	0.7	ID182	0.4	ID242	0.3
ID003	0.8	ID063	0.3	ID123	0.7	ID183	0.7	ID243	0.2
ID004	0.5	ID064	0.4	ID124	0.6	ID184	0.4	ID244	0.8
ID005	0.6	ID065	0.6	ID125	0.2	ID185	0.3	ID245	0.65
ID006	0.8	ID066	0.4	ID126	0.8	ID186	0.9	ID246	0.8
ID007	0.4	ID067	0.5	ID127	0.4	ID187	0.4	ID247	0.15
ID008	0.2	ID068	0.5	ID128	0.7	ID188	0.3	ID248	NA
ID009	0.3	ID069	0.7	ID129	0.7	ID189	0.5	ID249	0.7
ID010	0.3	ID070	0.2	ID130	0.3	ID190	0.3	ID250	0.35
ID011	0.7	ID071	0.2	ID131	0.6	ID191	0.6	ID251	0.7
ID012	0.3	ID072	0.3	ID132	0.4	ID192	0.5	ID252	0.45
ID013	0.1	ID073	0.1	ID133	0.5	ID193	0.7	ID253	0.8
ID014	0.1	ID074	0.5	ID134	0.6	ID194	0.4	ID254	0.6
ID015	0.4	ID075	0.4	ID135	0.2	ID195	0.8	ID255	0.3
ID016	0.3	ID076	0.3	ID136	0.5	ID196	NA	ID256	0.9
ID017	0.2	ID077	0.1	ID137	0.1	ID197	0.5	ID257	0.35
ID018	0.5	ID078	0.4	ID138	0.1	ID198	0.4	ID258	0.5
ID019	0.6	ID079	0.3	ID139	0.5	ID199	0.15	ID259	0.25
ID020	0.3	ID080	0.7	ID140	0.2	ID200	0.5	ID260	0.1
ID021	0.4	ID081	0.4	ID141	0.3	ID201	0.5	ID261	0.2
ID022	0.4	ID082	0.4	ID142	0.5	ID202	0.3	ID262	0.15
ID023	0.4	ID083	0.5	ID143	0.5	ID203	0.7	ID263	0.7
ID024	0.5	ID084	0.75	ID144	0.7	ID204	0.4	ID264	0.1
ID025	0.3	ID085	0.6	ID145	0.6	ID205	0.2	ID265	0.3
ID026	0.15	ID086	0.5	ID146	0.1	ID206	0.5	ID266	0.7
ID027	0.5	ID087	0.3	ID147	0.3	ID207	0.8	ID267	0.8
ID028	0.3	ID088	0.8	ID148	0.5	ID208	0.4	ID268	0.5
ID029	0.6	ID089	0.25	ID149	0.5	ID209	0.11	ID269	0.3
ID030	0.4	ID090	0.7	ID150	0.2	ID210	0.2	ID270	0.7
ID031	0.75	ID091	0.5	ID151	0.8	ID211	0.6	ID271	0.8
ID032	0.55	ID092	0.3	ID152	0.7	ID212	0.2	ID272	0.6
ID033	0.8	ID093	0.3	ID153	0.7	ID213	0.1	ID273	0.3
ID034	0.8	ID094	0.7	ID154	0.35	ID214	0.6	ID274	0.6
ID035	0.6	ID095	0.7	ID155	0.5	ID215	0.3	ID275	0.25
ID036	0.5	ID096	0.3	ID156	0.5	ID216	0.25	ID276	0.5
ID037	0.2	ID097	0.2	ID157	0.5	ID217	0.3	ID277	0.3
ID038	0.5	ID098	0.6	ID158	0.8	ID218	0.45	ID278	0.5
ID039	0.6	ID099	0.4	ID159	0.3	ID219	0.2	ID279	0.6
ID040	0.9	ID100	0.7	ID160	0.7	ID220	0.1	ID280	0.7
ID041	0.7	ID101	0.3	ID161	0.3	ID221	0.3	ID281	0.4
ID042	0.2	ID102	0.4	ID162	0.6	ID222	0.2	ID282	0.6
ID043	0.6	ID103	0.5	ID163	0.7	ID223	0.2	ID283	0.8

ID044	0.3	ID104	0.55	ID164	0.5	ID224	0.2	ID284	0.8
ID045	0.6	ID105	0.3	ID165	0.4	ID225	0.1	ID285	0.2
ID046	0.6	ID106	0.55	ID166	0.4	ID226	0.1	ID286	0.3
ID047	0.6	ID107	0.7	ID167	0.6	ID227	0.2	ID287	0.6
ID048	0.15	ID108	0.2	ID168	0.4	ID228	0.1	ID288	0.1
ID049	0.1	ID109	0.3	ID169	0.4	ID229	0.5	ID289	0.1
ID050	0.3	ID110	0.1	ID170	0.7	ID230	0.4	ID290	0.8
ID051	0.2	ID111	0.4	ID171	0.8	ID231	0.7	ID291	0.6
ID052	0.15	ID112	0.4	ID172	0.2	ID232	0.35	ID292	0.3
ID053	0.8	ID113	0.4	ID173	0.8	ID233	0.3	ID293	0.4
ID054	0.1	ID114	0.5	ID174	0.3	ID234	0.7		
ID055	0.3	ID115	0.4	ID175	0.7	ID235	0.3		
ID056	0.4	ID116	0.6	ID176	0.4	ID236	0.6		
ID057	0.5	ID117	0.5	ID177	0.3	ID237	0.8		
ID058	0.8	ID118	0.5	ID178	0.8	ID238	0.5		
ID059	0.4	ID119	0.9	ID179	0.2	ID239	0.5		
ID060	0.8	ID120	0.3	ID180	0.4	ID240	0.6		

Supplemental Table 2: List of genes included in the custom SureSelectXT panel (Agilent). Probes covering the full-length coding region of all 67 genes were included and subsequently used for SNV detection. Previously published in Hoyer et al.[1]

Gene	Gene	Gene	Gene	Gene	Gene
ACVR1B	*CDKN2A*	*KMT2C*	*PALB2*	*RBM6*	*SMARCA4*
ACVR2A	*EGF*	*KMT2D*	*PBRM1*	*RNF43*	*TGFBR1*
APC	*EGFR*	*KRAS*	*PIK3CA*	*ROBO1*	*TGFBR2*
ARID1A	*ERBB2*	*MAP2K4*	*PIK3CG*	*ROBO2*	*TLE4*
ARID2	*ERBB3*	*MAP2K7*	*PIK3R5*	*RPS6KA2*	*TP53*
ATM	*ErbB4*	*MAPT*	*PLCG2*	*SETD2*	*TP53BP2*
BCORL1	*FBXW7*	*MARK2*	*PLXNB2*	*SF3A1*	*U2AF1*
BRAF	*GATA6*	*MSH2*	*PMS2*	*SF3B1*	
BRCA1	*GLI3*	*MYC*	*PREX2*	*SLIT2*	
BRCA2	*GNAS*	*MYCBP2*	*PRKCG*	*SMAD3*	
CALD1	*HUWE1*	*NF2*	*PRSS1*	*SMAD4*	
CAMK2B	*KDM6A*	*NRG1*	*RBM10*	*SMARCA2*	

Supplemental Table 3: List of genes targeted by specific probes that are included in the PDAC custom NGS panel. 3-6 probes per genes, covering SNPs that are located before, behind and within the gene of interest were used for subsequent CNA detection. Previously published in Hoyer et al.[1]

Gene	Gene	Gene	Gene	Gene	Gene
ACVR1B	*LG2*	*KMT2A*	*NOV*	*ROBO1*	*CDKN2B*
AKT1	*CDH1*	*KMT2C*	*PALB2*	*ROBO2*	*ERBB3*
ARID1A	*CDK6*	*KMT2D*	*PBRM1*	*RPA1*	*NCOR1*
ARID1B	*CDKN2A*	*KRAS*	*PDCD1*	*SETD2*	*NCOR*
ATM	*EGFR*	*MAP2K4*	*PDGFRA*	*SF3B1*	*KDM6A*
BRAF	*ERBB2*	*MET*	*PIK3CA*	*SLIT2*	*RBM10*
BRCA1	*FGFR1*	*MIB1*	*PIK3R1*	*SMAD4*	*BCORL1*
BRCA2	*FGFR2*	*MLH1*	*PIK3R3*	*SMARCA2*	
CASP8	*FGFR3*	*MSH2*	*PREX2*	*SMARCA4*	
CCND2	*GATA6*	*MYB*	*PTEN*	*SOX9*	
CCNE1	*JAK2*	*MYC*	*RB1*	*STK11*	
CD274	*KIT*	*NOTCH1*	*RNF43*	*TGFBR2*	

Supplemental Table 4: 1086 SNVs detected by targeted sequencing in 293 PDAC patients. Previously published in Hoyer et al.[1]

See additional excel file

Supplemental Table 5: 4157 CNAs detected by targeted sequencing and MLPA analyses in 283 PDAC patients. Previously published in Hoyer et al.[1]

See additional excel file

Supplemental Table 6: List of pathways affected by either SNV or CNA in genes included in the PDAC panel. Each patient with a mutation or CNA in one of the genes, associated to the respective pathway is counted as affected by the pathway. Pathway affiliations according to KEGG and Bailey *et al*.

Function	Patients	% patients	Genes included
MAPK Pathway	288	98%	BRAF, FGFR1, FGFR2, FGFR3, KIT, KRAS, MAP2K4, MAP2K7, MAPT, MET, PDGFRA, RPS6KA2
Cell Cycle	270	92%	CCND2, CCNE1, CDK6, CDKN2A, CDKN2B, MYC, RB1, TP53, TP53BP2
TGFBeta Signalling	226	77%	ACVR1B, ACVR2A, SMAD3, SMAD4, TGFBR1, TGFBR2
PIK3-/AKT Pathway	190	65%	AKT1, MYB, PIK3CA, PIK3CG, PIK3R1, PIK3R3, PIK3R5, PREX2, PTEN, STK11
DNA Repair	175	60%	ATM, BRCA1, BRCA2, MLH1, MSH2, PALB2, PMS2, RPA1
Others	172	59%	CALD1, CASP8, CDH1, GATA6, GLI3, GNAS, HUWE1, MARK2, PLCG2, PLXNB2, PRKCG, PRSS1, SOX9
Chromatin - SWI/SNF	155	53%	ARID1A, ARID1B, ARID2, PBRM1, SMARCA2, SMARCA4
Chromatin	145	49%	KDM6A, KMT2A, KMT2C, KMT2D, NCOR1, SETD2, SMARCB1
NOTCH Signaling	141	48%	BCORL1, FBXW7, MIB1, NF2, NOTCH1, NOV
EGFR Signalling	128	44%	EGF, EGFR, ERBB2, ERBB3, ErbB4, NRG1
Immun Supression	118	40%	CD274, JAK2, PDCD1, PDCD1LG2
WNT Signalling	117	40%	APC, CAMK2B, RNF43, TLE4
RNA Processing	105	36%	RBM10, RBM6, SF3A1, SF3B1, U2AF1
ROBO SLIT	88	30%	MYCBP2, ROBO1, ROBO2, SLIT2

Supplemental Table 7: **List of *TP53* mutations with functional consequences (activating vs inactivating)[232].** Previously published in Hoyer et al.[1]

See additional excel file

Supplemental Table 8: Clinical baseline characteristic in all NMF cluster. For each characteristic both the total number of patient in each category and their percentage of the total patient count per cluster is given.

Characersitics		Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
		(n = 11)	(n = 29)	(n = 121)	(n = 50)	(n = 69)
Age, years						
	Median	68	63	64	64,5	64
	range	46 - 75	45-80	35-82	44-80	24-76
Sex						
	male - no. (%)	8 (73%)	18 (62%)	62 (51%)	22 (44%)	43 (62%)
	female - no. (%)	3 (27%)	11 (38%)	59 (49%)	28 (56%)	26 (38%)
Arm						
	Gemcitabine - no. (%)	6 (55%)	15 (52%)	58 (48%)	21 (42%)	41 (59%)
	Gemcitabine + Erlotnib - no. (%)	5 (45%)	14 (48%)	63 (53%)	29 (58%)	28 (41%)
Karnofsky						
	60 - no. (%)	0 (0%)	1 (3%)	0 (0%)	0 (0%)	0 (0%)
	70 - no. (%)	1 (9%)	2 (7%)	2 (2%)	2 (4%)	2 (3%)
	80 - no. (%)	3 (27%)	5 (17%)	33 (27%)	16 (32%)	14 (20%)
	90 - no. (%)	3 (27%)	12 (41%)	47 (39%)	16 (32%)	28 (41%)
	100 - no. (%)	4 (36%)	9 (31%)	39 (32%)	16 (32%)	25 (36%)
Grading						
	G1 - no. (%)	1 (9%)	0 (0%)	1 (1%)	2 (4%)	2 (3%)
	G2 - no. (%)	5 (45%)	14 (48%)	77 (64%)	35 (70%)	38 (55%)
	G3 - no. (%)	5 (45%)	13 (45%)	38 (31%)	12 (24%)	27 (39%)
	unknown - no. (%)	0 (0%)	2 (7%)	5 (4%)	1 (2%)	2 (3%)
T-Stage						
	T1 - no. (%)	1 (9%)	0 (0%)	0 (0%)	5 (10%)	2 (3%)
	T2 - no. (%)	1 (9%)	4 (14%)	11 (9%)	4 (8%)	9 (13%)
	T3 - no. (%)	9 (82%)	25 (86%)	107 (88%)	41 (82%)	58 (84%)
	T4 - no. (%)	0 (0%)	0 (0%)	3 (2%)	0 (0%)	0 (0%)
N-Stage						
	N0 - no. (%)	6 (55%)	9 (31%)	42 (35%)	16 (32%)	26 (38%)
	N1 - no. (%)	5 (45%)	20 (69%)	79 (65%)	34 (68%)	43 (62%)
Postoperative CA 19-9, kU/L						
	Median (range)	13 (1-1976)	31 (2-1000)	19 (1-2573)	11 (1-405)	12 (1-77)
	<= 100 - no. (%)	9 (82%)	25 (86%)	99 (82%)	34 (68%)	44 (64%)
	101-500 - no. (%)	1 (9%)	1 (3%)	10 (8%)	7 (14%)	9 (13%)
	> 500 - no. (%)	1 (9%)	1 (3%)	4 (3%)	4 (8%)	1 (1%)
	unknown - no. (%)	0 (0%)	2 (7%)	8 (7%)	5 (10%)	15 (22%)

Supplemental Table 9: Overview of differentially expressed genes in *SMAD4*^{alt} patients. All genes with adjusted p-value (BY.p.value) < 0.05 are shown. Modified from Hoyer et al.[1]

	fold change	Log2 fold change	std error (log2)	Lower confidence limit (linear)	Upper confidence limit (linear)	BY.p.value
SMAD4	-1,2932489	-0,371	0,0579	0,715	0,836	0,00000351
MLLT3	-1,3669875	-0,451	0,0959	0,642	0,833	0,00939
H3F3A	-1,2561424	-0,329	0,0721	0,722	0,878	0,0117
COL1A2	-1,5230887	-0,607	0,139	0,544	0,793	0,0165
GNAS	-1,2763288	-0,352	0,081	0,702	0,875	0,0165
PLD1	-1,271913	-0,347	0,0814	0,704	0,878	0,0165
BAD	-1,2544022	-0,327	0,0779	0,717	0,886	0,0165
NOTCH1	-1,1974787	-0,26	0,0613	0,768	0,907	0,0165
RAC1	1,23627526	0,306	0,0724	1,12	1,36	0,0165
MECOM	1,3425725	0,425	0,102	1,17	1,54	0,0165
NFKB1	1,40834423	0,494	0,119	1,2	1,65	0,0165
DTX4	1,4171574	0,503	0,122	1,2	1,67	0,0174
BRIP1	1,64490014	0,718	0,175	1,3	2,09	0,0174
SETBP1	-1,2807599	-0,357	0,0878	0,693	0,88	0,0182
CREBBP	1,53368266	0,617	0,151	1,25	1,88	0,0182
HHEX	-1,3764956	-0,461	0,115	0,621	0,849	0,0201
ASXL1	-1,1447242	-0,195	0,0486	0,818	0,933	0,0201
UBB	-1,3268451	-0,408	0,102	0,656	0,866	0,0203
MAPK9	-1,1883831	-0,249	0,0626	0,773	0,916	0,0203
IL20RB	1,90659709	0,931	0,235	1,39	2,62	0,0207
RAF1	-1,2058078	-0,27	0,0685	0,756	0,91	0,021
SMARCB1	-1,3745887	-0,459	0,118	0,62	0,853	0,0235
FZD7	-1,3158545	-0,396	0,103	0,661	0,873	0,0262
ALKBH3	1,25266444	0,325	0,0851	1,12	1,41	0,0292
CHEK2	1,43097652	0,517	0,136	1,19	1,72	0,0292
EZH2	1,3103934	0,39	0,103	1,14	1,51	0,0296
KIT	1,54756499	0,63	0,166	1,23	1,94	0,0296
FANCG	-1,3407126	-0,423	0,112	0,64	0,869	0,0312
LIFR	-1,3425725	-0,425	0,116	0,636	0,871	0,0421
IKBKB	-1,2474656	-0,319	0,0876	0,712	0,903	0,0455

Eigenständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig verfasst habe und sämtliche Quellen, einschließlich Internetquellen, die unverändert oder abgewandelt wiedergegeben werden, insbesondere Quellen für Texte, Grafiken, Tabellen und Bilder, als solche kenntlich gemacht habe.

Ich versichere, dass ich die vorliegende Abschlussarbeit noch nicht für andere Prüfungen eingereicht habe.

Mir ist bekannt, dass bei Verstößen gegen diese Grundsätze ein Verfahren wegen Täuschungsversuchs bzw. Täuschung gemäß der fachspezifischen Prüfungsordnung und/oder der Fächerübergreifenden Satzung zur Regelung von Zulassung, Studium und Prüfung der Humboldt-Universität zu Berlin (ZSP-HU) eingeleitet wird.

Berlin, 10.04.2021